

STAT

Conclusion

Crop Productivity and Meteorological Factors (Statistical
Investigations)

Gosplanizdat, Moscow, 1949

STAT

STAT

A CONTRIBUTION TO THE QUESTION OF FORECASTING THE CHARACTER
OF THE SPRING AND SUMMER WEATHER

THE FUNDAMENTAL THESES OF THE INVESTIGATION

(The author urgently recommends that the following system be strictly followed in reading this work: to guide oneself by "the fundamental theses of the investigation", in which are set forth in brief the contents of the work and the theses advanced by it, and then to supplement each thesis by the study of the corresponding part of the work. For this purpose, each point of the "fundamental theses" has been supplied with an indication of the corresponding page numbers of the work. This simultaneous use of theses and main body of the work will considerably facilitate the task of becoming acquainted with the investigation.)

I. This work establishes the existence of empirical regularities between certain indices of late autumn, winter and early spring weather, on the one hand, and the course of the spring and summer weather, on the other.

In the work herein presented the author has set himself the following object: to discover such indices in the march of the autumn, winter and early spring weather as would serve as signals of the character of the coming spring and summer period with respect to its humidity or aridity. Since the course of the weather during this spring and summer period, and the degree of its humidity or aridity, are principally reflected in the fluctuations of crop yields from year to year, the search for a method of correlating the parallelism between the yield fluctuations and the meteorological

indices of the periods preceding the spring and summer months was taken as the foundation.

The investigation has been based on the materials of the 33-year period from 1883 to 1915 during which the crop statistics work of the former Central Statistical Committee functioned (see pages 47-48 of the Introduction for details).

II. The author establishes the following basic symptoms which signalize the character of the forthcoming weather during the spring and summer period: (1) the date of the onset of spring in the Central Agricultural Region, as determined by the commencement of the unfolding of the leaves according to the observations of many years' standing made at the meteorological station of B. Fenino (observations of I. A. Pul'man); (2) the peculiarities in the movement of the daily mean temperatures over ten-day periods from October to April inclusive; (3) the difference between the precipitation in the southwest and the northeast of the European part of the USSR during the period preceding the spring, namely from November to April inclusive.

Let us consider each of these criteria separately.

1. The time of onset of spring in the Central Agricultural Region. The author arrives at the following thesis: the early onset of spring is an unfavorable symptom for the forthcoming spring and summer weather, while its late onset is a favorable sign. Further analysis showed that the time of the onset of spring serves as a signal in this direction not only for the forthcoming year, but also for the following year; i.e. the time of onset of spring in 1882, 1883, 1884, etc., down to 1914, was compared with the crop yield (by which

the character of the weather was reflected) in 1883, 1884, 1885, etc., down to 1915. Both symptoms, however, each taken separately, are only of slight forecasting value. Their combined use increases the accuracy of a forecast, but only to an insignificant extent (for details see pages 52-55). The prognostic significance of this index was verified over the 12-year period from 1920 to 1931, which was not included in the main investigation (which was from 1883 to 1915), and was completely confirmed thereby (see Chapter VII, pages 114, 117).

2. The peculiarities of temperature movements from October to April inclusive.

(a) The generally known law of temperature movements from autumn through winter into spring, consisting of their fall towards the winter and their subsequent rise from the winter towards the spring, may be expressed by a parabola of the second order. This law, however, is realized in a different way for each year. In some years the parabola that represents the movement of the temperatures may be steep, in other years gently sloping. In the former case we have a sharp fall in temperature from autumn to winter and a sharp rise from winter to spring, while in the latter case these contrasts are smoothed out.

These individual peculiarities of each year in the underlying dynamics of the temperature -- from autumn through winter to the spring -- proved to be signals of the humidity or dryness of the forthcoming spring and summer weather (insofar as these factors were reflected in the prerevolutionary harvests from 1883 to 1915). A steeper course of the temperature represents an unfavorable symptom, while the smoothing out of the dynamic contrasts is, on the contrary, a favorable symptom. It was, moreover, noted that the same kind of

parallelism between the fluctuations of the crop yields (in which the march of the spring and summer weather is reflected) and the peculiarities of the temperature dynamics from autumn through winter to spring was also observed with respect to the autumn-winter-spring of the preceding agricultural year (in a manner entirely analogous to the first criterion, the onset of spring). The simultaneous use of the index for two years does not, however, yield high correlation coefficients (see section 6, pages 55-65 for details).

(b) Both criteria (the onset of spring in the Central Agricultural Region, and the dynamics of temperature from autumn through winter to spring, in both cases for two years) are each individually only weak indicators of the forthcoming spring and summer weather, but their joint use is already enough significantly to increase the accuracy of the forecast (see section 7, pages 65-70 for details).

(c) Section 8 (pages 70-73) concerns an important question: to what territory does the effectiveness of the indications on the forthcoming weather extend (insofar as that weather was reflected in the fluctuations of the prerevolutionary crop yields)? Within the present (1940 -- The Editor) boundaries of the republics, oblasts and kray^s in the European part of the USSR, the joint use of these two indicators appears most effective for the continuous territorial bloc that embraces the northern part of the Ukraine, Kursk and Voronezh Oblasts, the south and east of Moscow Oblast, Ivanovo, Gor'kiy, Kirov, Sverdlovsk, Chelyabinsk, Kuibyshev, Orenburg, Saratov and Stalingrad Oblasts, and the northern part of the Azov-Black Sea Kray. For the south of the Ukraine, and as well as for the south of the Azov Black

Sea Kray and North Caucasus, the significance of these indicator-criteria is apparently considerably more attenuated.

(d) The indicator-criterion: steepness or gentleness of the slope of the parabola characterizing the march of temperatures from autumn through winter to spring will be analyzed in detail and made more precise in the sequel. This detailed analysis and refinement of a second indicator-criterion, i.e. of the steep or gentle slope of the parabola (the first criterion, as we have seen, is the date of the onset of spring), reduces the system of indicator-criteria to one of the second order, thereby considerably increasing the effectiveness of the forecasts. We now proceed with the exposition of this system. The first type of parabola (with a steep slope) indicates an augmented difference between the winter temperature, on the one hand, and the autumn and spring temperatures on the other; the higher the autumn and spring temperatures, and the lower those in winter, the more sharply expressed is the first type of curve. On the other hand, the second type of curve corresponds to a cold autumn and spring and a warm winter. If we take the correlations between the crop yields and the ten-day temperatures from October to April inclusive we shall be convinced that high temperatures in October, March and April and low temperatures in winter correspond in general to diminished yields, while the opposite relation between the temperatures (cold October, March and April, and warm winter) on the other hand represents a favorable sign. The extreme ten-day periods are the most symptomatic of all: the first in October and the third in April -- and also the median period, i.e. the third in January. In addition, the abruptness of the transition from autumn to winter occupies a special position: the greater the difference

between the last two warm ten-day periods of autumn and the first two cold periods of winter, the more unfavorable the forecast. Thus the second sign (the peculiarities of the march of temperatures from autumn through winter to spring) is detailed and broken down into four more elementary signs: (1) the mean temperature of the first ten-day period of October (symbol O_1); (2) the mean temperature of the third ten-day period of January (symbol J_3); (3) the temperature of the third ten-day period of April (symbol A_3); (4) the temperature difference between the last two warm ten-day periods and the first two cold ten-day periods (symbol aP_w) which characterizes the transition from autumn to winter. To these four elementary indices two more, relating to the prior agricultural year, are added, namely: (5) the first ten days of October of the prior agricultural year (symbol O_{-1}), and (6) the temperature of the third ten days of April of the prior agricultural year (symbol A_{-1}).

From these six elementary indices an integral index, denoted by the symbol T_6 , is constructed. Its forecasting significance is measured by the correlation coefficient -0.754 (Ul'yanovsk) and -0.723 (Kazan'). The combined use of the two basic indices: (1) the date of the onset of spring; and (2) the character of the movement of autumn, winter and early spring temperatures in the form of the index T_6 -- leads to an increase of forecast accuracy to 0.811 (see Chapter II, pages 74-80, and Chapter III, pages 80-90).

(e) The index T_6 , which characterizes the peculiarities of the dynamics of autumn, winter and early spring temperatures, has been verified against the material of years not included within the scope of the present investigation, namely, 1872-1882, and, what is especially important, on material of 1920-1931, and its forecast

significance has been confirmed, as has also been that of the first basic criterion (the commencement of the onset of spring in the Central Chernozem Region) (see Chapter VII, Section 18, pages 109-114).

3. We pass now to the third basic index: the difference in the rainfall from October to April inclusive in the northeastern region of the European part of the USSR and that rainfall in its southwestern region. The original work was done at the stations of Ul'yankovsk -- Kazan', on the one hand, (representing the northeast), and Kiev -- Odessa, on the other hand, (representing the southwest). It developed that the greater this difference was in favor of the northeast, the more reason there was to expect favorable spring and summer weather (insofar as the latter was reflected in the fluctuations of crop yields between 1883 and 1915). The forecasting significance of this indicator was equal to plus 0.502. Detailed analysis of the investigation showed, however, that all the six months from November to April inclusive were not active in this connection, but only three: November, January and April. Taking this additional detail into consideration increased the correlation coefficient to plus 0.635 (see Chapter IV, pages 91-93). This indicator was subsequently verified on the basis of materials of 22 eastern stations and 16 southwestern stations, and for these the correlation coefficient proved to be somewhat higher (+0.635). The use of these 38 stations, together with 33 supplementary ones, made it possible to fix schematic boundaries for the northeast and for the southwest. The northeast region proved to have approximately the following delineation: its northern boundary runs along the line Moletovsk-Kirov (Vyatka); thence southwestward through Gor'kiy,

Voronezh and Ostrogozhsk stations; thence northeastward through Penaa, Ul'yanovsk, Polibino, Ufa, Zlatoust and back to Molotovsk.

The approximate eastern boundary of the southwest region follows the line of the Dnepr and Desna, thence turns west and runs through Vasil'vich to Pinsk. From Pinsk its northern boundary runs to Warsaw, Berlin and Trier. From Trier the line turns and runs down to Milan, Rome and Sassari, whence it turns back through Yugoslavia (Khvar), Rumania (Sulina), to the Crimean peninsula (or the mouths of the Danube), where it meets the southwestern region of the USSR (see Chapter IV, Section 14, pages 93-97).

Unfortunately I had no opportunity to verify the forecasting significance of this criterion from the materials of years other than those included in this investigation, as was done with the first and second criteria. For this reason I advance this third criterion only as a hypothesis which, though very probable, still lacks adequate verification.

III. (a) In Chapter V an integral index is constructed from the nine indicators of the forthcoming spring and summer weather that have been considered above. These 9 indicators are first resolved into 11 elementary indices. The forecasting value of this integral index, thus combining the elementary indicators, is measured by a coefficient of correlation equal to 0.867 (see Chapter V, pages 97-99).

(b) This generalized index could unfortunately not be verified by the author from materials of other years than those comprised in the investigation, since he had no data on one of the important indicators, divided into two elementary indices, namely:

on the difference in rainfall (for November, January and April) between the northeast and southwest; but an internal verification was made by dividing the 33-year period into two special periods, one from 1883 to 1899, and the second from 1900 to 1915. The regression equation was calculated separately for each of these periods. The regression equation of the first period was then applied as a forecasting equation to the second period, and conversely: the regression equation for the second period was used as a forecasting equation for the first. The indices thus calculated according to the "extraneous" equation were compared with the actual crop yields, whose fluctuations reflected the march of the spring and summer weather. The coincidence of these values proved to be significant, and was measured by coefficients of correlation ranging from 0.885 to 0.825 (see Chapter VI, pages 100-105).

This internal verification, which nevertheless could not entirely take the place of an external verification, in any event does allow us to accept as a very probably hypothesis the view that the significant forecasting value of the integral index summates the forecasting values of all the elementary criteria it includes.

(c) It is repeatedly emphasized in the present investigation that the crop-yield fluctuations from 1883 to 1915 are regarded in this work as a very sensitive indicator on which is impressed the march of the spring and summer weather and its degree of aridity or humidity, and that these prerevolutionary crop yields are considered here exclusively from that point of view. In Chapter VII an attempt is made to dispense with this indirect indicator and to pass to the direct data on the spring and summer weather. This gave entirely satisfactory results (see Chapter VII, section 17, pages 105-109).

IV. In conclusion, the author maintains the necessity of continuing the work. Such continuation should be carried out in the following directions: (1) external verification for periods not included in the investigation proper. This has been done already to a considerable extent for the two basic criteria (early onset of spring and dynamics of temperature from autumn through winter into spring) but has not been done for the third basic indicator (the differences in the November, January and April rainfall between northeast and southwest); (2) transition from the indirect index of the course of the weather, furnished by the crop yield fluctuations from 1883 to 1915 to the direct meteorological indices; (3) increasing the precision of the definition of territory and time interval covered by the forecasting significance of each individual criterion. (pages 117-118).

INTRODUCTION

1. The results of the investigation herein presented are based on materials for 33 years (1883 to 1915), and disclosed empirical regularities between certain indices of late autumn, winter, and early spring weather on the one hand, and the course of the spring and summer weather on the other hand. These regularities appear to be purely empirical, and no physical explanation of them is given in the present work. The chain of causal dependences that connect the indices included in this investigation with the character of the spring and summer weather can be formulated only in a very general and abstract form, namely, as follows: The indices with which the investigation is concerned evidences and are symptoms of those processes taking place in the atmosphere of late summer and early spring, which are of decisive importance for the aridity of humidity of the weather during the spring and summer period.

Nevertheless such empirical regularities are of scientific importance even where the physical meaning of such causal dependencies still remain undiscovered, under the indispensable condition, however, that such dependencies, which are founded only on observation and experiment, actually exist and reflect phenomena actually occurring in reality. It is possible to become convinced of this latter condition only through practice. By the application of the observed relations and of the formulae (in which these relations find their mathematical expression) to the phenomenal of actual life, the degree of their justification is determined, which serves in the last analysis as a criterion of the extent to which the empirical regularities (and the formulae that express them) reflect reality.

Such a verification was made to some extent by the author, and yielded entirely satisfactory results. The author, however, does not consider the work of verification to have been completed, and therefore assigns to the conclusions of the investigation only the significance of a hypothesis (a very probable hypothesis but still hypothesis). Whether this hypothesis takes on the character of a proved fact, whether it receives partial confirmation, or is to be completely rejected, depends on the results of further verification.

It goes without saying that if this result should be favorable the urgent necessity of the theoretical working out of the regularities established will then be in order, together with the discovery of the meteorological, soil, and phytophysiological processes that underlie such regularities.

2. The author set the following goal for the present work: To find those indices in the course of autumn, winter, and early spring weather which would serve as signals of the character with respect to humidity or aridity of the forthcoming spring and summer period. Inasmuch as the fluctuation in the crop yields from year to year, before the revolution, mainly reflected the course of the spring and summer weather (its degree of humidity or aridity) the investigation was based on the search by the correlation method for a parallelism between the fluctuations of the yields and the meteorological indices of the periods preceding the spring and summer months. The data on yields have been taken from my work "Movement of Crop Yields of Grain Crops in European Russia, 1883 to 1915" (page 56) and are presented in the table on page 49.

The first column of the table gives the indices of harvest per desyatin for all crops, according to data of the former Central Statistical Committee. In the second column, this data, aligned by the method of least squares, reflect the growth of productivity throughout the entire period. The third column shows the departures of the actual yields from the corresponding "norms" of productivity for each year (column 1 minus column 2). Column 4 expresses these fluctuations as percentage ratios. Thus the last two series characterized the fluctuations of the yield, from which is excluded (to the maximum extent possible with the available material) the influence of the growth of agricultural technology, and by reason of which they reflect far better the action of the meteorological conditions, than the series of direct crop yields data, in which their dynamics is a function of factors of various orders: on the one hand meteorological, on the other hand agricultural. In the investigation we make use of the latter series, which characterizes in percent the variability of the crop yields.

MOVEMENT OF CEREAL CROP YIELDS IN 50 SUBJECTIONS OF FORMER
EUROPEAN RUSSIA, 1883-1915

YEAR	YIELD OF CEREALS	YIELD NORMS	DEPARTURES OF YIELDS FROM NORMS	
	IN PUDS PER DESYATIN	IN PUDS	IN PUDS	IN PERCENT
1	2	3	4	5
1883	35.2	35.1	+ 0.1	+ 0.5
1884	38.2	35.6	+ 2.5	+ 7.0
1885	32.2	36.1	- 3.9	- 10.8
1886	36.6	36.6	+ 0.0	+ 0.0

1	2	3	4	5
1887	42.2	37.0	+ 5.4	+ 14.6
1888	40.9	37.5	+ 3.4	+ 9.1
1889	31.1	38.0	- 6.9	- 18.2
1890	35.1	38.5	- 2.8	- 7.3
1891	27.7	39.0	- 11.3	- 29.0
1892	34.3	39.5	- 5.2	- 13.2
1893	48.7	40.0	+ 8.7	+ 21.8
1894	49.4	40.5	+ 8.9	+ 22.0
1895	45.0	41.0	+ 4.0	+ 9.8
1896	43.5	41.5	+ 2.0	+ 4.8
1897	36.0	42.0	- 6.0	- 14.3
1898	41.9	42.4	- 0.5	- 1.2
1899	47.2	42.9	+ 4.3	+ 10.0
1900	44.5	43.4	+ 1.1	+ 2.5
1901	37.8	43.9	- 6.1	- 13.9
1902	51.2	44.4	+ 6.8	+ 15.3
1903	46.7	44.9	+ 1.8	+ 4.0
1904	54.0	45.4	+ 8.6	+ 18.9
1905	42.5	45.9	- 3.4	- 7.4
1906	35.3	46.4	- 11.1	- 23.9
1907	42.2	46.9	- 4.7	- 10.0
1908	43.3	47.3	- 4.0	- 8.5
1909	55.3	47.8	+ 7.5	+ 15.7
1910	51.4	48.3	+ 3.1	+ 6.4
1911	39.9	48.8	- 8.9	- 18.2

1	2	3	4	5
1912	53.2	49.3	+ 3.9	+ 7.9
1913	58.3	49.8	+ 8.5	+ 17.1
1914	44.5	50.3	- 5.8	- 11.5
1915	55.8	50.3	+ 5.5	+ 10.9

3. As noted above, the investigation was based on the comparison of various kinds of meteorological indices of winter and early spring with the fluctuations of the yields. This was done with the aim of determining the symptoms of the character of spring and summer weather to be expected. It is understandable that it was only possible to use the data on yields because the character of the weather from 1883 to 1915 with respect to its aridity or humidity, was also mainly responsible for the fluctuation of crop yields during the prerevolutionary period. The series selected by us relates on the whole to the territory of the former 50 guberniyas of European Russia. (The crop yield material of the former TsSK does not go back as far as 1883 for the other territories that compose the USSR. Even for the territories and guberniya of North Caucasus, the TsSK began to collect yield data only in 1892, and still later for the other territories.)

During the second stage of the work, this territory was more defined as follows: It was proved that the forecasting criteria selected were mainly effective for the territory of the resultant belts. The following oblasts are included in this more precisely bounded territory: The northern part of the Ukraine (the southern part only weakly so)(Vinnitsa, Kamenets-Podol'sk, Zhitomir, Kiev, Poltava, Kharkov, and Donetsk Oblasts), Rostov Oblast, the Central

Chernozem oblasts (Kursk, Orlov, Voronezh, Tambov, Tula, Penza, Ryazan), the Volga oblasts (Stalingrad, Saratov, Kuibyshev, Ul'yans'k), the Chkalov, Chelyabinsk, Sverdlovsk, Kirov, Gor'kiy, and Ivanovo Oblasts, and the Bashkir, Tatar, Udmurt, Mari, Chuvash, Mordov AO. Since crop yield data was not available for Siberia and Kazakhstan a far back as 1883, they were not included in the investigation.

During the third stage of the work the general forecasting index constructed on the basis of the above operations was directly verified by comparison of the rainfall and temperatures for the three months of May, June and July on the precisely defined territory.

Finally, the last and concluding phase of the investigation -- the verification of the empirical regularities found, and of the equations expressing them, to see whether they corresponded with reality -- was only partially performed and cannot be considered completed.

4. In his use of the correlation method, the author encountered the following complications.

As is well known, the full coefficient of correlation of multiple regression, commonly denoted by the symbol R , includes the systematic error, which exaggerates its real value. This systematic error increases, other things being equal, with the number of variables that enter into the regression equation. At the limit, when the number of variables in the regression equation becomes equal to the number of their observations, the full coefficient of correlation reaches unity, regardless of whether there is any dependence of the phenomenon y on the other phenomena x, c, t, u, v , etc., or not. In this limiting case, all of the values of y , calculated according

to the regression equation, coincide with the observed values. It is clear in what the danger of using the empirical correlation coefficient consists: It bears elements of illusion, mirage, and the results of the investigation represented by it have a more optimistic character than is justified by reality. It is specially important to eliminate the systematic error of the correlation coefficient of multiple regression in cases when empirical regularities are to be established which still have no physical explanation. The present work is an example. The following formula may be proposed as an approximation, but accurate enough for 33 observations:

In this formula R denotes the empirical correlation coefficient of multiple regression. R' the corrected correlation coefficient from which the systematic error has already been excluded; N is the number of observations processed by the method of least squares, and n is the number of variables in the regression equation. In the frequent case of a regression equation with two variables $y = a + bx$, R coincides in value with r_{xy} , and the corrected coefficient of correlation is equal to:

$$R' = \sqrt{1 - \frac{n-1}{N-2} (1 - R^2)}$$

The precise formula derived by Wishart for correcting the empirical coefficient of multiple correlation is complicated. After a few transformations it may be given the following form:

$$1 - R^2 = \frac{N-1}{N-n} (1 - R^2) \quad (1)$$

Where R is the empirical coefficient of multiple correlation, R^2 is the corrected correlation coefficient sought, N is the number of observations, n is the number of variables in the regression equations, and f is a certain function of the number of observations N and the value of the correlation coefficient such that

$$f = \frac{1}{2} R^2 + \frac{1}{2} R^4 + \frac{1}{2} R^6 + \frac{1}{2} R^8 + \dots$$

This series is divergent to infinity, and it can easily be seen that its divergence increases rapidly as the number of observations increases, and as R decreases.

We turn now to formula (1) and after simple transformations we find

$$R^2 = 1 - \frac{N-1}{N-n} \cdot \frac{1-R^2}{1+f}$$

The simplified formula for the corrected correlation coefficient, as we have seen, is of the following form:

$$R^2 = 1 - \frac{N-1}{N-n} (1 - R^2)$$

The exact formula is thus distinguished from the approximate formula by the denominator $1 + R^2$ under $1 - R^2$. But $1 + R^2 > 1$ in any case.

$$\text{Thus: } \frac{N-1}{N-R} (1-R) > \frac{N-1}{N-R} \frac{(1-R^2)}{(1+R^2)},$$

$$1 - \frac{N-1}{N-R} (1-R) < 1 - \frac{N-1}{N-R} \frac{(1-R^2)}{(1+R^2)},$$

$$\text{or } 'R < P < R$$

We thus arrive at a very important conclusion.

Calculation by the simplified formula gives a value for the correlation coefficient that is less than its real value, i.e. it introduces a larger correction than it should. There is no great harm in this since it is better to be somewhat over zealous in correcting the empirical coefficient R than to underestimate the extent of the systematic error.

The situation thus mathematically proven is expressed by the following formula:

$$'R < P < R$$

CHAPTER I

5. Everyone who has had to work on questions connected with crops, has involuntarily turned his attention to the fact that all of the major catastrophic crop failures caused by drought (in 1891, 1906, and 1921) have been preceded by an early spring. The early onset of spring was also frequently observed in years with less severe crop failures (for example, 1885, 1890, 1901, 1920). This symptom, however, has no decisive "prognostic" value, and drought and a poor harvest do not always follow an early spring, for instance the years 1913 and 1930 had good harvests and also early spring. The author commenced his work by verifying this approximate connection and determining the degree of its closeness.

The phenological observations at the Bogoroditskoye-Penino meteorological station were taken as indices of the time of onset of spring. (The B. Penino station is located in Kursk Oblast. The observations were conducted by I. A. Pul'man and include the entire period under study from 1883 to 1915 inclusive. The author was compelled to confine himself to the observations at this station, since no other equally long phenological series was available for other points. The observations were commenced by Pul'man in 1882, and have continued uninterruptedly down to the present.) These observations covered the time of the unfolding of leaves on trees (lilacs, birches, and elders). This sign is not one of the first signs of spring; on the contrary, it marks the end of the first period of spring and the beginning of the second period of spring, when, with the commencement of the activity of the chlorophyll granules, the plants enter the period of the intensified growth and development.

The 33-year period was divided into five groups of 6 or 7

years each. The first group includes the 6 years with the earliest onset of spring (17 to 26 April). The second includes the 7 years when the leaves unfolded from 27 April to 1 May. The third included the 7 years with dates between 1 May and 4 May. The fourth included the 7 years with dates from 4 May to 7 May. The final group included the 6 years with the latest onset of the second period of spring (unfolding of the leaves 7 May to 12 May). The mean crop yield was calculated for each group in percent below (-) or above (+) the norms.

The results of this classification are given in the following table:

TABLE 1

Classification of Years by Time of Opening of Leaves
on Trees at B. Fenino Station
(time given is in days since 1 April)

Departure of Crop Yields
in % above (+) or below (-)

6 years with earliest opening of leaves (17-27 days)	(-) 4.3
7 years with opening of leaves from 27-31 days	(-) 4.6
7 years with opening of leaves from 31-34 days	(+) 0.7
7 years with opening of leaves from 34-37 days	(+) 4.9
6 years with latest opening of leaves (37-42 days)	(+) 4.0

This table shows that the crop yields increase (but slightly) as we pass from years with early onset of spring to years with later onset; the first two groups with unfolding of leaves before 1 May have crop yields 4.3 percent and 4.6 percent under the average respectively; the third group coincides almost exactly with the average; and, finally, the last two groups with leaves opening on and after 4 May are characterized by crop yields 4.9 percent and 4.0

percent above the average.

The correlation coefficient between the dates of leaf opening and the size of the harvests is + 0.263 and confirms the existence of a slight parallelism between the time of onset of spring and the crop yields.

Further analysis shows that the time of onset of spring signals in the same direction and with the same force (i.e. weakly) not only the forthcoming crop yield, but also that of the following year.

On comparison of the dates of leaf opening in 1882, 1883, 1884, etc., down to 1914 with the respective harvests of 1883, 1884, 1885, etc., down to 1915, we prepared another table analogous to the first one, giving the following data:

TABLE 2

Classification of Years by Dates of Opening of Leaves on Trees in the Preceding Year (at B. Fenino Station)	Departure of Crop Yields in % above (+) or below (-)
6 years with earliest opening of leaves (17-27 days)	(-) 8.4
7 years with opening of leaves from 27-31 days	(-) 3.5
7 years with opening of leaves from 31-34 days	(+) 6.2
7 years with opening of leaves from 34-37 days	(+) 2.5
6 years with latest opening of leaves (37-42 days)	(+) 1.0

It appears from this table that the average yield in the first group with the earliest unfolding of the leaves in the year preceding the harvest was 8.4 below the norm, in the following group, with a later unfolding of the leaves the yield was 3.5 percent below the norm

while the highest yield of all + 6.2 falls in the middle group, which forces us to assume the existence of a curvilinear correlation. (The author, however, was compelled to ignore this similarity, in view of the complexity of the question. This simplification of the problem is reflected in the results of the investigation in the direction of making the correlation coefficient smaller than in reality, not in exaggerating it. For this reason there can be no objection against the high coefficients that are obtained in the final results.)

The correlation coefficients between the crop yields and the date the leaves unfolded in the preceding year is expressed by the same value, 0.267. Thus the prediction of the crop yield is also weak according to this criterion.

The question arises: is not the significance of these indicators increased when they are used together?

In order to solve this question, we calculated a series of indices of the mean date of leaf unfolding for two years (for instance, 1882 to 1883, 1883 and 1884, 1884 and 1885, etc., down to 1914-1915) and Table 3 which is analogous to the first two tables, has been prepared according to this criterion:

TABLE 3

Classification of Years by Average Date of Unfolding of Leaves for Two Successive Years (at B. Penino)	Departure of the Crop Yields in % above (+) and below (-)
6 years with earliest unfolding of leaves (21-28.5 days)	(-) 8.6
7 years with unfolding of leaves from 28.5 to 30 days	(-) 3.5
7 years with unfolding of leaves between 30.5 and 33.5 days	(+) 0.5
7 years with unfolding of leaves from 33.5 to 36.0 days	(+) 5.4
6 years with latest unfolding of leaves (36-39.5 days)	(+) 5.8

This table gives the same picture of the variation of crop yields with the increasing lateness of spring as is given by the first two tables, with the character of its expression sharper, however. In the first table the sum of the departures of the crop yields from their mean for all 5 groups is 18.5 (4.3 + 4.6 + 0.7 + 4.9 + 4.1), in the second table this sum is 21.6 (8.4 + 3.5 + 6.2 + 2.5 + 1.0), while in the last table it reaches 23.8.

The correlation coefficient between crop yield and mean time of leaf unfolding for two years is + 0.355, which still more definitely marks a certain improvement in the forecasting of the forthcoming crop, but only a "certain" improvement, since the size of the correlation coefficient evidences the low significance of both criteria even when jointly used.

The parallelism noted is thus very probable. As we shall see further on, a final and positive answer is given when the index we have just considered is used to supplement the more effective and more important indicator to the establishment of which we now pass.

6. This section considers the peculiarities of the movement of the ten-day mean daily temperatures from October to April inclusive. The "normal" dynamics of temperature, which consists in its gradual fall from ten-day period to ten-day period, until the middle of January is reached, and in its likewise gradual rise from then to the end of April, is completely realized only in the aggregate for many years. As may be seen from the data of Table 4, even the 39-40 year period at Ul'yanovsk meteorological station gives some slight irregularities in the march of temperatures from autumn to winter and from winter to spring.

In individual years, however, this regularity is seriously impaired by the zigzag course of the temperatures, but their fundamental dynamics, conditioned by cosmic causes (the revolution of the earth around the sun and the inclination of the earth's axis to the plane of the ecliptic) is, of course, observed annually.

To represent it, recourse must be had to the mathematical methods of trend fitting, the aim of which is to reveal the "law" of the movement of temperature from autumn through winter to the spring, thus eliminating all these disturbances in its actual march. A parabola of the second order is a function entirely suitable for this purpose: it is able to reflect the fall in temperature that then gives place to a rise, without taking in the disturbances in the fundamental dynamics, since it possesses only a single maximum or minimum, as the case may be. (Parabolas of the third and higher orders, which have more than one maximum or one minimum, may "catch" these disturbances, which is why their use in this case involves risk.)

We see from the example of the trend fitting to the 39-40 year observations at Ul'yanovsk station, which yield an almost regular dynamics of temperature for the period taken, that a parabola of the second order reproduces it satisfactorily: the sum of the unsquared deviations between the observations and the calculated temperature indices is 17.7, which gives us a mean deviation of $\frac{17.7}{21} = 0.84$ degree; the sum of the squares of these non-coincidences is 20.7 degrees and the standard deviation $\sigma = 0.99$ degree; finally, the coefficient of multiple correlation, corresponding to the regression equation, is 0.989, or almost unity, which shows that 85 percent of the entire dynamics (fundamental and incidental) is

covered by the equation of the parabola. (As a matter of fact, this coincidence between the second-order parabola and the actual march of the temperature should really be greater than this, since the actual mean temperatures over the 39-40 year period for which the comparison is made, still reproduces the dynamics with a certain amount of distortion: it is enough to point out that from the third decade in December to the second decade in February inclusive there is no regular fall and rise in the temperature indices from one decade to the next, and, of course, during this interval the fitted indices reproduce the actual processes better.)

TABLE 4

MEAN DIURNAL TEMPERATURES AT UL'YANOVSK STATION, 1876/77 TO 1914/15

MONTHS	DECADES	MEAN DIURNAL		FITTED INDICES	DEVIATIONS	
		TEMPERATURES			[5]	x
	[1]	[2]	[3]	[4]		[6]
October	1st	+ 6.3		+ 8.1	- 1.8	- 10
	2nd	+ 3.8		+ 4.2	- 0.4	- 9
	3rd	+ 0.5		+ 0.7	- 0.2	- 8
November	1st	- 2.1		- 2.4	+ 0.3	- 7
	2nd	- 4.1		- 5.2	+ 1.1	- 6
	3rd	- 5.6		- 7.5	+ 1.9	- 5
December	1st	- 7.7		- 9.3	+ 1.6	- 4
	2nd	- 9.9		- 10.8	+ 0.9	- 3
	3rd	- 12.6		- 11.9	+ 0.7	- 2
January	1st	- 12.4		- 12.6	+ 0.2	- 1
	2nd	- 14.3		- 12.9	- 1.4	0
	3rd	- 13.6		- 12.6	- 1.0	+ 1
February	1st	- 11.6		- 12.2	+ 0.6	+ 2

	[1]	[2]	[3]	[4]	[5]	[6]
March		2nd	- 12.3	- 11.2	- 1.1	+ 3
		3rd	- 11.0	- 9.9	- 1.1	+ 4
		1st	- 7.7	- 8.1	+ 0.4	+ 5
		2nd	- 7.1	- 5.9	- 1.2	+ 6
		3rd	- 3.3	- 3.4	+ 0.1	+ 7
		1st	+ 0.1	- 0.4	+ 0.5	+ 8
April		2nd	+ 3.3	+ 3.0	+ 0.3	+ 9
		3rd	+ 7.7	+ 6.8	+ 0.9	+ 10

Equation of parabola:

$$y = - 12.8576 - 0.0652 x + 0.2031 x^2; \text{ (see Dia ram 1)}$$

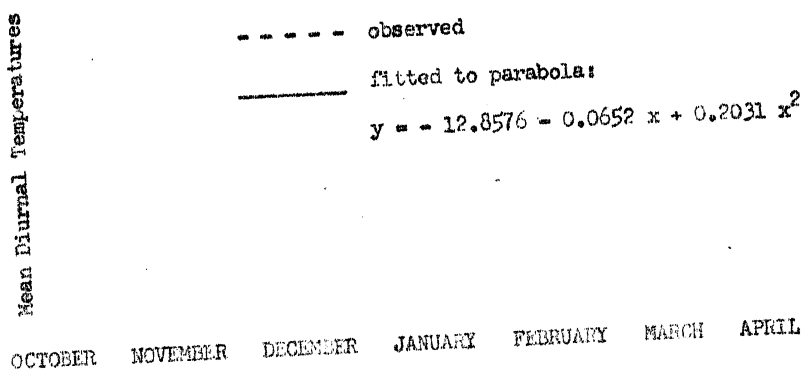
$$R = 0.989 \quad (1)$$

We dwell on one element that characterizes the parabola, namely, on the degree of steepness or gentleness with which it slopes. Its angle of inclination at its initial or final points (i.e. at $x = (-) 10$ and $x = (+) 10$) may serve as a criterion of this. The matter thus resolves itself into a very simple mathematical task: to find the angular coefficients to the parabola at the points $x = (-) 10$ and $x = (+) 10$. As is well known, the angular coefficient is measured by the tangent of the angle between tangent and x-axis.

The value of these angular coefficients for the parabola presented above is $(-) 4.13$ for the autumn, and $(+) 4.00$ for the spring. There is almost no difference between their absolute values, since the curve is almost symmetrical about the y-axis. The signs $(+)$ and $(-)$ for angular coefficients do not refer to the magnitudes of the angles of intersection, and only indicate their differing directions, and accordingly are devoid of significance here. In our

subsequent discussion we shall employ only the absolute magnitudes of the angular coefficients.

Diagram 1. Mean diurnal temperatures by decades over 39-40 years
(Ul'yanovsk station)



Just as the 10-day temperatures from October to April over the 39-40 year period were fitted to a parabola of the second order, so we proceed with respect to each individual year.

For the sake of graphic illustration we present three examples of the determination of the basic dynamics of the temperature from autumn through winter to spring: 1885/86, 1890/91 and 1901/02 (see Table 5 on page 58).

Table 6 contains for each year the equations of the parabola, the angular coefficients and the complete correlation coefficients corresponding to each parabola (see page 59).

It is clear from these tables that the parabolas, in eliminating the disturbances in the regular march of temperature from autumn through winter to spring, not only reveal the general factors common to each year but also reflect the individual peculiarities of the basic process.

TABLE 5

MEAN DIURNAL TEMPERATURES AT UL'YANOVSK STATION IN 1885/86, 1890/91 AND 1901/02

Month	10-Day Period	Number of 10-Day Period	x	1885/86 Temperature		1890/91 Temperature		1901/02 Temperature	
				Observed	Fitted	Observed	Fitted	Observed	Fitted
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
October	1st	1	- 10	+ 5.8	+ 8.4	+ 6.5	+ 8.0	+ 4.5	+ 5.4
	2nd	2	- 9	+ 4.1	+ 4.5	+ 1.4	+ 4.7	+ 0.7	+ 2.7
	3rd	3	- 8	+ 1.8	+ 0.7	+ 2.5	- 0.8	- 0.5	- 0.1
November	1st	4	- 7	- 4.2	- 2.7	+ 2.3	- 4.4	- 3.4	- 2.4
	2nd	5	- 6	- 5.7	- 5.4	- 8.0	- 7.6	+ 0.3	- 4.4
	3rd	6	- 5	- 12.0	- 8.0	- 17.6	- 10.2	- 4.0	- 6.2
December	1st	7	- 4	- 3.8	- 10.0	- 7.4	- 12.5	- 9.9	- 7.8
	2nd	8	- 3	- 7.6	- 11.6	- 14.2	- 14.0	- 7.5	- 8.8
	3rd	9	- 2	- 12.2	- 12.8	- 17.0	- 15.0	- 12.1	- 9.6
January	1st	10	- 1	- 6.6	- 13.7	- 13.9	- 15.5	- 2.1	- 10.1
	2nd	11	0	- 10.2	- 13.8	- 22.7	- 15.6	- 15.8	- 10.4
	3rd	12	+ 1	- 18.9	- 13.6	- 26.6	- 15.2	- 11.3	- 10.3

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
February	1st	13	+ 2	- 18.9	- 13.2	- 1.2	- 14.2	- 12.2	- 10.0	
	2nd	14	+ 3	- 10.5	- 12.2	- 11.9	- 12.7	- 8.0	- 9.3	
	3rd	15	+ 4	- 19.7	- 10.9	- 10.5	- 10.7	- 9.3	- 8.4	
March	1st	16	+ 5	- 10.5	- 9.2	- 2.5	- 8.2	- 8.4	- 7.3	
	2nd	17	+ 6	- 8.0	- 7.0	+ 0.1	- 5.2	- 8.9	- 5.8	
	3rd	18	+ 7	- 4.6	- 4.2	+ 1.6	- 1.5	- 2.8	- 4.0	
April	1st	19	+ 8	+ 2.7	- 1.4	+ 1.2	+ 2.4	+ 0.9	- 1.9	
	2nd	20	+ 9	+ 4.7	+ 2.1	+ 4.8	+ 7.0	- 0.2	+ 0.5	
	3rd	21	+ 10	+ 6.1	+ 6.0	+ 8.3	+ 12.0	+ 4.0	+ 3.2	

Regression Equations

1885/86:	$y = 13.8128$	$-0.1188x$	$0.2102 x$	; R	0.875
1890/91:	$y = 15.6446$	$-0.2062 x$	$0.2560 x$	; R	0.874
1901/02:	$y = 10.4776$	$-0.1083 x$	$0.1481 x$	; R	0.865

Angular coefficients:

	Autumn (with $x = 10$)	Spring (with $x = 10$)
1885/86	4.32	4.09
1890/91	4.93	5.34
1901/02	3.07	2.85

TABLE 6

FITTING OF THE MEAN DIURNAL TEMPERATURES FOR 10-DAY PERIODS FROM OCTOBER TO APRIL INCLUSIVE

Years	Equations of Regression			Angular Coefficients		R
				Autumn [5]	Spring [6]	
[1]	[2]	[3]	[4]			[7]
1881/82	$y = -10.9231$	$-0.0382 x$	$+0.1505 x^2$	3.04	2.96	-
1882/83	$y = -13.4847$	$+0.0330 x$	$+0.2014 x^2$	4.00	4.06	0.841
1883/84	$y = -12.6204$	$-0.3738 x$	$+0.1876 x^2$	4.13	3.38	0.902
1884/85	$y = -10.7088$	$-0.2484 x$	$+0.1792 x^2$	3.83	3.34	0.918
1885/86	$y = -13.8128$	$-0.1188 x$	$+0.2162 x^2$	4.32	4.09	0.875
1886/87	$y = -9.9549$	$-0.0683 x$	$+0.1503 x^2$	3.09	2.92	0.800
1887/88	$y = -12.3538$	$-0.0208 x$	$+0.2174 x^2$	4.37	4.33	0.880
1888/89	$y = -16.1569$	$-0.1843 x$	$+0.2504 x^2$	5.19	4.82	0.875
1889/90	$y = -12.0013$	$+0.2113 x$	$+0.2209 x^2$	4.21	4.63	0.936
1890/91	$y = -15.6446$	$+0.2062 x$	$+0.2568 x^2$	4.93	5.34	0.874
1891/92	$y = -12.5732$	$-0.0791 x$	$+0.1878 x^2$	3.84	3.68	0.766
1892/93	$y = -15.9199$	$+0.0575 x$	$+0.2443 x^2$	4.83	4.94	0.847
1893/94	$y = -10.6278$	$-0.2546 x$	$+0.1783 x^2$	3.82	3.31	0.928

[1]	[2]	[3]	[4]	[5]	[6]	[7]
1894/95	$y = -12.1737$	$-0.0255 x$	$+0.1732 x^2$	3.49	3.44	0.878
1895/96	$y = -16.6111$	$-0.3838 x$	$+0.2526 x^2$	5.44	4.67	0.962
1896/97	$y = -15.3782$	$-0.1609 x$	$+0.2547 x^2$	5.22	4.93	0.953
1897/98	$y = -14.9782$	$-0.2760 x$	$+0.1953 x^2$	4.18	3.63	0.861
1898/99	$y = -9.7239$	$-0.0016 x$	$+0.1426 x^2$	2.85	2.85	0.768
1899/1900	$y = -16.7188$	$-0.2900 x$	$+0.2726 x^2$	5.74	5.16	0.949
1900/01	$y = -11.7876$	$+0.1253 x$	$+0.2306 x^2$	4.49	4.74	0.933
1901/02	$y = -10.4776$	$-0.1083 x$	$+0.1481 x^2$	3.07	2.85	0.865
1902/03	$y = -11.1585$	$+0.3670 x$	$+0.1780 x^2$	3.19	3.93	0.865
1903/04	$y = -12.0359$	$-0.0262 x$	$+0.1707 x^2$	3.44	3.39	0.810
1904/05	$y = -12.5855$	$-0.1100 x$	$+0.2143 x^2$	4.40	4.18	0.913
1905/06	$y = -11.8854$	$-0.1287 x$	$+0.2517 x^2$	5.16	4.91	0.937
1906/07	$y = -13.8203$	$+0.0035 x$	$+0.2148 x^2$	4.29	4.30	0.896
1907/08	$y = -15.9055$	$-0.0978 x$	$+0.2207 x^2$	4.51	4.32	0.863
1908/09	$y = -14.5072$	$+0.0869 x$	$+0.1951 x^2$	3.82	3.99	0.943

[1]	[2]	[3]	[4]	[5]	[6]	[7]
1909/10	$y = -12.7724$	$- 0.1248 x$	$+ 0.2450 x^2$	4.97	4.89	0.950
1910/11	$y = - 14.0234$	$- 0.0352 x$	$+ 0.2092 x^2$	4.22	4.15	0.857
1911/12	$y = - 12.4498$	$- 0.1500 x$	$+ 0.1906 x^2$	3.96	3.66	0.849
1912/13	$y = - 11.2231$	$+ 0.0982 x$	$+ 0.1897 x^2$	3.70	3.89	0.824
1913/14	$y = - 7.0438$	$- 0.0774 x$	$+ 0.1165 x^2$	2.41	2.25	0.793
1914/15	$y = - 10.1549$	$+ 0.1170 x$	$+ 0.1590 x^2$	3.06	3.30	0.825

We pass now to the analysis of these tables.

First of all we remark that the minimum points of almost all years belong to the median ten-day period of the interval under study, namely, to the second ten-day period in January. There are two exceptions in all to this rule, neither of them very important at that. In 1883/84 the minimum of the parabola fell in the third ten-day period of January, and in 1902/03 it fell in the first ten-day period of the same month. (It is easy to convince oneself of this. All that need be done is to take the first derivative of the regression equation, equate it to zero and then find x .) This speaks for the view that the basic process in each year is markedly symmetrical. In the second place, our attention is called to the considerable fluctuations in the coefficient R . As has already been said, this index indicates to what extent the observed temperatures differ from those calculated from the parabola. In our case they may be used as indices of the degree of disturbance in the basic dynamics of temperature from autumn through winter to spring: the lower the coefficient R , the more strongly is the regular march of temperature disturbed; and the higher the value of R , on the other hand, the more closely does their actual movement correspond to the basic process in the year in question. In the third place, the angular coefficients characterize the degree of steepness or gentleness of the parabola's slope.

[See next page for Diagrams 2 and 3]

A sharper fall in the temperature from autumn to winter and a steeper rise from winter to spring corresponds to an increase in the angular coefficients; on the other hand, a smoother dynamics

We pass now to the analysis of these tables.

First of all we remark that the minimum points of almost all years belong to the median ten-day period of the interval under study, namely, to the second ten-day period in January. There are two exceptions in all to this rule, neither of them very important at that. In 1883/84 the minimum of the parabola fell in the third ten-day period of January, and in 1902/03 it fell in the first ten-day period of the same month. (It is easy to convince oneself of this. All that need be done is to take the first derivative of the regression equation, equate it to zero and then find x .) This speaks for the view that the basic process in each year is markedly symmetrical. In the second place, our attention is called to the considerable fluctuations in the coefficient R . As has already been said, this index indicates to what extent the observed temperatures differ from those calculated from the parabola. In our case they may be used as indices of the degree of disturbance in the basic dynamics of temperature from autumn through winter to spring: the lower the coefficient R , the more strongly is the regular march of temperature disturbed; and the higher the value of R , on the other hand, the more closely does their actual movement correspond to the basic process in the year in question. In the third place, the angular coefficients characterize the degree of steepness or gentleness of the parabola's slope.

[See next page for Diagrams 2 and 3]

A sharper fall in the temperature from autumn to winter and a steeper rise from winter to spring corresponds to an increase in the angular coefficients; on the other hand, a smoother dynamics

Diagram 2. Mean diurnal temperatures by ten-day periods, 1890/91
(Ul'yanovsk Station)

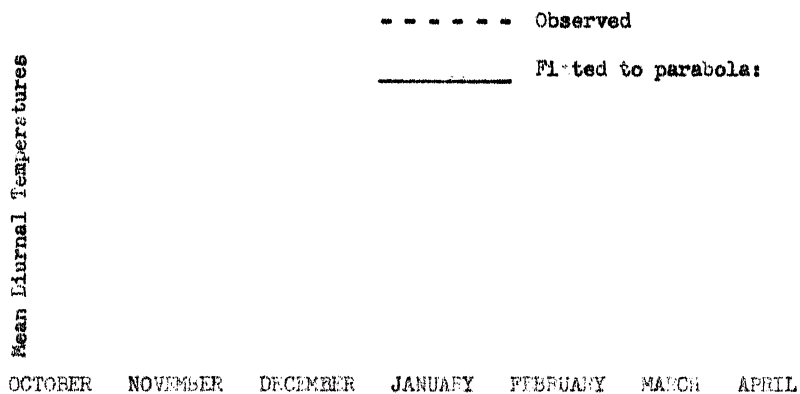
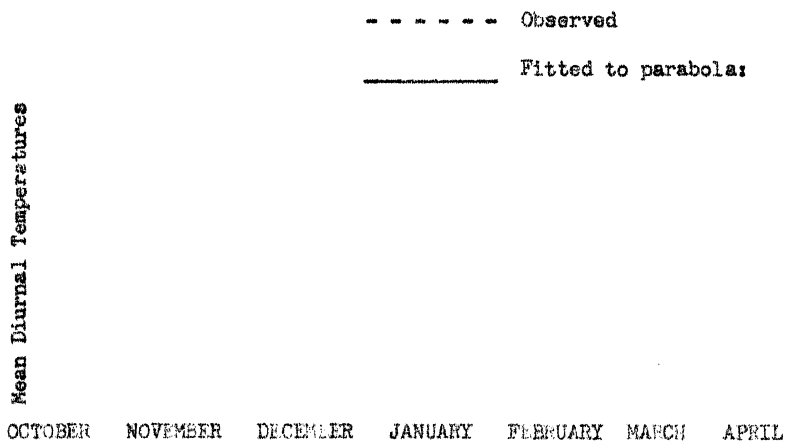


Diagram 3



of the temperature in the period involved corresponds to lower angular coefficients. The varying course of the basic process is graphically demonstrated by three diagrams, which show the temperature regimes of 1890/91 (steep parabola and high angular coefficients), 1901/02 (gentle parabola and low angular coefficients) and 1885/86 (intermediate type parabola and average size angular coefficients, cf. diagrams 2, 3 and 4).

We now consider whether the individual peculiarities of each year in the basic temperature dynamics -- peculiarities reflected in the sharper or smoother course of the process -- signalize anything with respect to the forthcoming crop yield, and consequently also with respect to the weather.

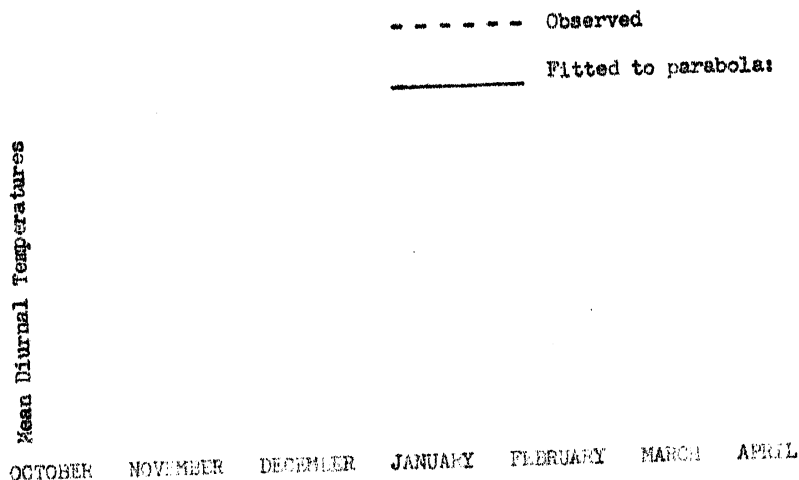
We draw up a table analogous to Tables 1, 2 and 3. We take the sum of the autumn and spring angular coefficients as the class criterion.

We obtain the following indices as a result:

TABLE 7

CLASSIFICATION OF YEARS BY SUM OF AUTUMN AND SPRING ANGULAR COEFFICIENTS	DEPARTURE OF
	CROP YIELDS
	in % over (+) and under (-)
6 years with highest values for angular coefficients (10.90-10.01)	(-) 13.0
7 years with angular coefficients from 9.80 to 8.59	(-) 0.3
7 years with angular coefficients from 8.58 to 7.62	(-) 0.5
7 years with angular coefficients from 7.59 to 6.93	(+) 5.1
6 years with lowest angular coefficients (6.83 to 4.66)	(+) 9.7

Diagram 4. Mean diurnal temperatures by ten-day periods, 1885/86
(Ul'yanovsk Station)



As may be seen from this table, the 6 years with highest angular coefficients yielded a harvest 13 percent below the average, the next two groups with lower angular coefficients are characterized by average crop yields, the fourth group, with coefficients from 7.59 to 6.93, has a crop yield 5.1 percent higher than average; finally, the group of years with lowest angular coefficients yielded the largest harvests, amounting to 9.7 percent above the average. Thus, to judge by this treatment, the individual peculiarities in the basic dynamics of temperature shown by each year forecast the forthcoming crop yield. A steeper course of the temperature is an unfavorable sign, while a smoother dynamics is, on the contrary, a favorable sign. However, even this indicator, though it forecasts more clearly than that established by us at the beginning of the chapter (the date of onset of spring, as determined by the unfolding of the leaves), is still a weak sign: the coefficient of correlation between the crop

yield and the angular coefficients is - 0.417.

We now pose the question whether these peculiarities in the basic dynamics of temperature -- from autumn through winter to spring -- in the prior agricultural year are not likewise symptoms of one crop yield or another. We took up the same question with respect to the first sign, and now, by analogy, put it with respect to the second. To decide it we prepare a table using the same angular coefficients as the class criterion, but moving them back by one year. Thus we compare the crop yield of 1883 with the angular coefficients of the same "October-April" period of 1881/82, the crop yield of 1884 with the angular coefficients for the same period of 1882/83, etc., and, finally, the crop yield of 1915 with the angular coefficients for 1913/14. As a result we obtain the following table:

TABLE 8

CLASSIFICATION OF YEARS BY SUM OF AUTUMN AND SPRING ANGULAR COEFFICIENTS FOR THE PRIOR YEAR	DEPARTURE OF CROP YIELD IN % OVER (+) AND UNDER (-)
6 years with highest values for angular coefficients of prior year (10.90 - 10.01)	(-) 10.0
7 years with angular coefficients 9.80 to 8.59	(-) 3.0
7 years with angular coefficients 8.58 to 7.62	(+) 4.2
7 years with angular coefficients 7.59 to 6.93	(+) 4.5
6 years with lowest angular coefficients (6.83 - 4.66)	(+) 3.3

The character of the parallelism is the same as that in the preceding table: years with crop yields under the group average correspond to the highest angular coefficients, while, in general,

low angular coefficients accompany crop yields over the average. (Apparently (just as with the first indicator) we have to do here with a curvilinear correlation: the third and fourth groups show almost no difference in the size of the yield, while the fourth group is even a little lower. Unfortunately, in view of the complexity of processing the materials in dealing with curvilinearity, it was necessary to ignore this similarity (see footnote at beginning of chapter).)

The correlation coefficient between crop yields and the angular coefficients for the prior year is - 0.369. This is not a high value, and no substantial forecasting results can be obtained from this indicator.

Joint use of the indicators for both years, however, does give clearer predictions. Table 9 shows the results of grouping by the average angular coefficients for the two successive years:

TABLE 9

CLASSIFICATION BY MEAN SUM OF ANGULAR COORDINATES FOR TWO YEARS	DEPARTURE OF	
	CROP YIELD	
	IN % OVER (+) AND UNDER (-)	
6 years with highest value of angular coefficients (10.1 to 9.3)	(-) 17.7	
7 years with angular coefficients 9.3 to 8.6	(-) 3.3	
7 years with angular coefficients 8.5 to 7.8	(+) 8.6	
7 years with angular coefficients 7.7 to 7.0	(+) 8.5	
6 years with lowest angular coefficients (7.0 to 5.5)	(+) 7.0	

The first group, with the highest angular coefficient for

two years, gives an average yield 17.7 percent below the norm, the next group is characterized by a yield somewhat below normal, and the remaining three groups with angular coefficients of 8.6 and lower have yields above normal. The parallelism between the angular coefficients for two successive years and the yield is very noticeable and is expressed far more sharply than in all the preceding analogous tables. This is especially marked in relation to the poor harvests. (Curvilinearity likewise appears in Table 9. For the third time I must express my regret that I was compelled to ignore this resemblance.)

All of the 10 years, with average angular coefficients for the two years of 8.9 and higher, are characterized, without exception, by crop yields that are under average, bad and very bad. These years are the following: 1897 (coefficient 10.1; crop yield (-) 14.3); 1901 (10.1 and (-) 13.9 respectively); 1891 (9.6 and (-) 29.0); 1889 (9.4 and (-) 18.2); 1890 (9.4 and (-) 7.3); 1906 (9.3 and (-) 23.9); 1907 (9.3 and (-) 10.0); 1911 (9.1 and (-) 18.2); 1898 (9.0 and (-) 1.2); 1892 (8.9 and (-) 13.2). The underlined years in this list are those with the worst crop failures in the period from 1883 to 1915. We add that only 4 of the remaining 23 years are marked by reduced crop yields (1905, 1908, 1914, 1885, with yields 7.4 to 11.5 percent below normal). There is thus an indubitable parallelism between the individual peculiarities in the dynamics of temperature from autumn through winter to spring and the crop yields. These peculiarities, as we have repeatedly said, consist in the sharp or gradual march of the basic temperature process.

The correlation coefficient between crop yields and the angular coefficients taken for two successive years also speaks in

favor of the indubitable character of this parallelism, for it is
- 0.526.

7. We have thus noted two signs that serve to forecast the forthcoming crop yield: the date of onset of spring in the central agricultural region (or, to put it more accurately, the second half of spring, since it is reflected in the unfolding of leaves on the trees), and the sharp or mild character of the basic temperature dynamics -- from autumn through winter to spring -- in the region of the middle Volga and the lower Kama. (Although we established this only for the single station of Ul'yanovsk, still, in the first place, temperature observations even at a single point are enough to well characterize its march for a very large region, and, in the second place, supplementary work at "Kazan" station, as we shall see below, fully confirmed the conclusions for Ul'yanovsk. This is why we are able to speak here of the middle Volga region and the Kama lowlands. Other regions will be mentioned in the following chapters.) Both these signs manifest themselves in the course of two successive years. Each of these signs, taken separately (especially the first one) is of a force that is most inadequate for any forecast. Using the method of least squares, we calculate the regression equation that defines the parallelism between the crop yield and these indices:

$$y = 13.6364 - 8.1135 X_c + 1.6425 B_c; R = 0.732 \quad (2)$$

The symbol y denotes the crop yield in percent departures from the trend line (see table on page 49); X_c is the mean sum of angular coefficients for two years; B_c is the mean date of unfolding of leaves on trees, likewise for two years (at B. Fenino Station).

The coefficient of multiple correlation suddenly jumps to

0.732. Even after correcting for its systematic error in multiple regression, by using the formula given in the Introduction, (with this correction of the empirical value of R bending it over somewhat towards the other side), this coefficient still does not drop below 0.704. (The actual correlation coefficient P by the formula given on page 52 lies within the limits $0.704 < P < 0.732$.)

The meaning of a regression equation that yields a correlation coefficient of this size stands out with great clarity and precision if we remember that it half answers the question. If we succeeded in finding one other sign, or several other signs, not connected correlatively with these two, but yielding by their own regression equation a correlation coefficient similarly equal to 0.70, then we should have completely solved the problem of forecasting. The combination of these new signs with those we had already should yield a coefficient of multiple correlation equal to unity, which would evidence complete coincidence of forecasts with reality. Such an ideal solution of the problem is, of course, still an unattainable goal at the present time. But this is why a regression equation like this one is of no little importance, since it advances the question with great power, and means a long step forward.

We still dwell on the coefficients of partial correlation, which disclose the significance of one factor on condition that the other be eliminated.

$$B_c \bar{r}_y, X_c = -0.685 \text{ and } X_c \bar{r}_y, B_c = +0.598$$

Table 10 shows all the indices, correlation coefficients and regression equations that have been discussed in this chapter. The assumed crop yields have been calculated from the numerical values

assumed by these indices in the several years in question, and the departures of the actual yields from these assumed yields have been given. In a word, this table recapitulates all that has been set forth in this chapter.

TABLE 10

YEAR	y	X	X ₋₁	X ₀	B	B ₋₁	B ₀	1Y	y - 1Y
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
1883	+ 0.5	8.1	6.0	7.0	30	31	30.5	+ 6.9	- 6.4
1884	+ 7.0	7.5	8.1	7.8	27	30	28.5	- 2.8	+ 9.8
1885	- 10.8	7.2	7.5	7.3	26	27	26.5	- 2.1	- 8.7
1886	0.0	8.4	7.2	7.8	42	26	34.0	+ 6.2	- 6.2
1887	+ 14.6	6.0	8.4	7.2	37	42	39.5	+ 20.1	- 5.5
1888	+ 9.1	8.7	6.0	7.4	28	37	32.5	+ 7.0	+ 2.1
1889	- 18.2	10.0	8.7	9.4	34	28	31.0	- 11.7	- 6.5
1890	- 7.3	8.8	10.0	9.4	23	34	28.5	- 15.8	+ 8.5
1891	- 21.0	10.3	8.8	9.6	31	23	27.0	- 19.9	- 9.1
1892	- 13.2	7.5	10.3	8.9	35	31	33.0	- 4.4	- 8.8
1893	+ 21.8	9.8	7.5	8.6	40	35	37.5	+ 5.5	+ 16.3
1894	+ 22.0	7.1	9.8	8.4	32	40	36.0	+ 4.6	+ 17.4
1895	+ 9.8	6.9	7.1	7.0	34	32	33.0	+ 11.0	- 1.2
1896	+ 4.8	10.1	6.9	8.5	42	34	38.0	+ 7.1	- 2.3
1897	- 4.3	10.2	10.1	10.1	32	42	37.0	- 7.5	- 6.8
1898	- 1.2	7.8	10.2	9.0	29	32	30.5	- 9.3	+ 8.1
1899	+ 10.0	5.7	7.8	6.8	32	29	30.5	+ 8.6	+ 1.4
1900	+ 2.5	10.9	5.7	8.3	35	32	33.5	+ 1.3	+ 1.2

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
1901	- 13.9	9.2	10.9	10.1	37	35	36.0	- 9.2	- 4.7	
1902	+ 15.3	5.9	9.2	7.6	36	37	36.5	+ 11.9	+ 3.4	
1903	+ 4.0	7.1	5.9	6.5	31	36	33.5	+ 15.9	- 11.9	
1904	+ 18.9	6.8	7.1	7.0	38	31	34.5	+ 13.5	+ 5.0	
1905	- 7.4	8.6	6.8	7.7	37	38	37.5	+ 12.8	- 20.2	
1906	- 23.9	10.1	8.6	9.3	23	37	30.0	- 12.5	- 11.4	
1907	- 10.0	8.6	10.1	9.3	35	23	29.0	- 14.2	+ 4.2	
1908	- 8.5	8.8	8.6	8.7	34	35	34.5	- 0.3	- 8.2	
1909	+ 15.7	7.8	8.8	8.3	36	34	35.0	+ 3.8	+ 11.9	
1910	+ 6.4	9.8	7.8	8.8	27	36	31.5	- 6.0	+ 12.4	
1911	- 18.2	8.4	9.8	9.1	29	27	28.0	- 14.2	- 4.0	
1912	+ 7.9	7.6	8.4	8.0	35	29	32.0	+ 1.3	+ 6.6	
1913	+ 17.1	7.6	7.6	7.6	25	35	30.0	+ 1.2	+ 15.9	
1914	- 11.5	4.7	7.6	6.1	17	25	31.0	- 2.2	- 9.3	
1915	+ 10.9	6.4	4.7	5.5	26	17	21.5	+ 4.3	+ 6.6	

From Table 10 we obtain:

Correlation Coefficients

Partial Correlation Coefficients

$$r_{y_1x} = - 0.417$$

$$B_c r_{y_1x_c} = - 0.685$$

$$r_{y_1x_1} = - 0.369$$

$$x_c r_{y_1B_c} = + 0.598$$

$$r_{y_1x_c} = - 0.526$$

$$r_{y_1B} = + 0.263$$

$$r_{y_1B_1} = + 0.283$$

$$r_{y_1B_c} = + 0.355$$

In this table, y is the crop yield, X is the sum of the angular coefficients (for autumn and spring), X_{-1} is the same for the prior year, X_c is the mean value of the same for two successive years $\frac{X+X_{-1}}{2}$; B is the date of leaf unfolding at B. Fenino station; B_{-1} is the same for the prior year; B_c is the mean value of the same for two successive years $\frac{B+B_{-1}}{2}$; y the assumed yield calculated from the regression equation:

$$R \text{ (corrected)} = 0.704$$

The days of leaf opening are counted from 1 April.

(To characterize the dynamics of the temperature from autumn through winter to spring, we have adopted the measure of steepness in the course of the parabola $t = a + bx + cx^2$, which is obtained by taking the sum of the absolute values of the angular coefficients (first derivatives) at the initial and end points of the segment of the parabola (in our case with $x = -10$ and $x = +10$). It is easy to convince oneself that c , the coefficient of x^2 , can replace the sum of the angular coefficients. To take the sum of the absolute values of autumn and spring angular coefficients is equivalent to subtracting the negative autumn coefficient algebraically from the positive spring coefficient. In the general formulae this gives

$$y - 10 = b - 20c \text{ and } 1y + 10 = b + 20c; 1y + 10 - 1y - 10 = 40c.$$

In this way we convince ourselves that the sum of the absolute values of the angular coefficients for autumn and spring is equal to c

multiplied by the constant coefficient (40), and this assures identity of the coefficients of correlation between yield and parameter c , on the one hand, and between yield and the sum of the angular coefficients, on the other. Similar results are likewise obtained for the regression equations, except that the coefficient of c is 40 times as large as that for X , since c itself is only 1/40th as large as X .

Diagram 5. Comparison of crop yields observed and theoretical yields calculated from regression equation:

$$y = 13.6364 - 8.1135 c + 1.6425 B_c$$

Departure of observed yields from the norms

Yields calculated from regression equation:

$$y = 13.6364 - 8.1135 c + 1.6425 B_c$$

The column "years" shows the crop years; the values of X are calculated from materials relating to two years (the autumn months and December of the past year, and January, February, March and April of the current year) and are entered by the respective years of the spring months. For instance, X calculated from data of October, November, December, 1885 and January, February, March and April, 1886, is entered on the line for 1886.

On the annexed Diagram 5, which illustrates the results of the investigation, the values of y calculated from regression equation (2) are plotted along the axis of abscissas, while the actual yields, in percent of the trend norms y , are plotted along the axis of ordinates. Each year of the 33 is denoted by the point having the corresponding values of y and y . For instance, 1891, with $y = -19.9$ and $y = -29.9$ is marked by the point measuring -19.9 along the axis of abscissas and -29.9 along the axis of ordinates.

The points entered are joined by lines to the axis of abscissas (the zero line of departures of the actual yields). This line graphically shows the departures of the yield above (+) or below (-) the trend line (see "Introduction" on the determination of these departures).

On examination of the data in Table 10, and of the annexed Diagram 5, we reach the following conclusions.

The criteria established for forecasting the forthcoming crop yield make no small contribution to the cause of forecasting. Assume that the course of the weather during the autumn, winter and spring, up to May, gives us no indication on the forthcoming harvest.

In this case only one answer can be given to the question as to what kind of crop yield can be expected: orientation toward its mean value assures the lowest probability of error. It is a different story if the meteorological processes during the preceding period give some signals on the harvest, and consequently also on the weather. In this case we are already enabled to dispense with the "average" and orient our forecast towards the changes of which the symptoms speak. The mathematical formulation of the effectiveness of these signals, as we have done, will give the quantitative expression of these forecasts. Comparison of their results with the "forecasts" made without benefit of such orientation (i.e. according to the "averages") is a criterion of their relative accuracy. Let us see what the regression equation worked out by us gives us in this respect.

For this purpose we draw up a table of forecast errors resulting from the orientation to the average yields, and of those resulting from the orientation to the regression equation.

Size of Errors	Number of Errors	
	From the Orientation to the Average	From the Orientation to the Regression Equation (series $y - \bar{y}$)
20% and over	4	1
15% - 20%	6 19	3 8
10% - 15%	9	4
Under 10%	<u>14</u>	<u>25</u>
Total	33	33

Thus we see that orientation to the "average" gives a preponderance of considerable errors (10 percent and over), amounting to 19 cases out of 33, while the regression equation reduces the number of such errors to 8, which is a quarter of the cases observed. Further, we see from the data of Table 10 that 15 of the forecasts predicted a crop yield below the average (the minus sign before y), that in 13 out of these cases the actual yield coincided in sign (the minus sign in the y series as well) and only in two was there a discrepancy in sign (1884, +7.0 and 1910, +6.4). In 18 cases the regression equation predicted an increased crop yield (the plus sign in the y series), and only in one case out of these 18 did the sign fail to coincide with the actual yield (1905, when it was - 7.4). Exceptions are thus rarely encountered, and when they are, they are small. (All these three cases were under 10 percent). Finally it may be pointed out that the 12 extreme variants have no exceptions at all, with the under average yields giving better forecast results: all the 9 forecasts from - 17.5 to -19.9 correspond to crop yields with lower indices (1889, 1890, 1891, 1897, 1898, 1901, 1906, 1907, 1911), and on the opposite side there are only three variants and no exceptions (1887, 1903, 1904).

The regression equation set up by us with two criteria, contributes no small amount to forecasting; but it is, nevertheless, unqualifiedly inadequate.

Let us consider in some detail the most favorable of the indicators for the regression equation, namely the coincidence of the plus and minus signs between forecast and fact. (This means that a precise, quantitatively expressed forecast is not required

for the equation, but merely an indication as to whether the forthcoming crop yield will be above average or below it.)

As we have seen, a divergence was observed in only 3 cases out of 33. The frequency of coincidence was 0.91, and that of the contrary event 0.09, with a mean square error of 0.05

($\sqrt{0.05} = 0.22$).

We pass in conclusion from frequency to probability, and by the rule of the "three sigmas" we determine almost with certainty (999 to 1) that the probability corresponding to our frequency cannot fall below 0.76. The justification of forecasts of coincidence of plus or minus signs can be entirely guaranteed only to the extent of 3 cases out of 4, if the series of experiments is long enough. If more exacting demands are made with respect to quantitative coincidence of forecasts with actual yields, then it is clear from the data presented that large errors (over 10 percent) are possible fairly often, for approximately a quarter of the cases fall within their range, while the possibility even of very large errors (20 percent in one case) is not excluded (though it is true that such errors are rare).

Our regression equation is thus inadequate and under all circumstances requires improvement.

We leave the solution of this problem to the following chapter, and pass now to the question of the territorial significance of the indicator-criteria.

8. Up to now we have been establishing a parallelism between the indicators and the average crop yield in each year for

the 50 guberniyas of former European Russia. The question of the precise delimitation of the territory is of great interest and importance: whether the forecast is effective for all the regions of this territory, or whether there are regions for which it is negligible or even entirely invalid, while for other regions the significance of such indicators may be delineated with sufficient clarity and definition.

In order to solve this problem we correlate the index I_y obtained from the equation, and which reflects the joint effect of two symptoms (X_0 and B_0), with the crop yield in each of the 50 guberniya of former European Russia. The indices of yield were taken from the works cited in the Introduction to this paper.

As a result we obtained the following 50 correlation coefficients.

[See Table 11 on following page]

These data speak with entire distinctness for the regional significance of our indicators. The guberniyas marked by the sign * give correlation coefficients of at least 0.40 with the crop yield, and form a continuous bloc which is hatched on the chart. All the guberniyas in this region, without exception, have coefficients of correlation of at least 0.40. In the remaining guberniyas, which are located to the north and south of this region, the coefficients are considerably lower and, in general, of negligible size (except for the guberniyas of Yekaterinoslav, Minsk, Mogilev and Astrakhan, which immediately adjoin the region). Applying the same principle of continuity of territory, we are able to separate still another

TABLE 11

Former guberniya	Former r guberniya	Former r guberniya	Former r
Arkhangal'sk	-0.09 Kostroma	+0.22 * Samara	+0.56
Vologda	-0.05 Urodno	+0.25 * Saratov	+0.59
Olonets	-0.13 Minsk	+0.30 * Tambov	+0.61
Novgorod	-0.18 Mogilev	+0.32 * Ryazan	+0.58
Estonia	-0.00 Chernigov	+0.20 * Tula	+0.40
Livonia	-0.04 * Perm	+0.66 * Orlov	+0.43
Kurland	+0.20 * Orenburg	+0.45 * Kursk	+0.51
Koyzo	+0.20 * Ufa	+0.51 * Voronezh	+0.58
Vilna	+0.05 * Vyatka	+0.50 * Don Terr.	+0.48
Vitebsk	-0.10 * Kazan'	+0.66 * Kharkov	+0.56
Pskov	+0.24 * Nizhniy- Novgorod	+0.60 * Poltava	+0.50
Iver	+0.10 * Vladimir	+0.41 " Kiev	+0.42
Smolensk	+0.16 * Simbirsk	+0.65 * Podolia	+0.42
Kaluga	+0.20 * Penza	+0.55 Yekaterinos- lav	+0.39
Moscow	+0.21	Bessarabin	+0.13
Yaroslavl	+0.10	Kherson	+0.20
		Tavris	+0.22
		Astrakhan	+0.31

subregion from the main massif. This subregion is more heavily hatched on the chart and consists of the guberniyas with correlation coefficients of at least 0.50. Further use of the continuity principle does not lead to a more detailed division.

Thus, for example, if we try to delimit the region with correlation coefficients of at least 0.68, we find it impossible to form any such region, for the Perm guberniya, with a correlation coefficient of 0.66, is separated by the Vyatka Ufa guberniyas from Kazan' guberniya, which has the same coefficient, 0.66.

CHART I

CORRELATION COEFFICIENT BETWEEN CROP YIELDS AND THE COMBINED EFFECT OF SYMPTOMS X_0 AND B_0 .

- LEGEND: I: Guberniya with correlation coefficients under 0.49.
II. Guberniya with correlation coefficients 0.40 to 0.49.
III. Guberniya with correlation coefficients 0.50 and higher.

NOTE: Yekaterinoslav guberniya, with a coefficient of 0.39, is divided into two parts. The northern part is shown as belonging to the territory where the indicators are effective, and the southern part to the territory where they are only feebly effective.

Our indicators are thus not effective throughout the entire area of former European Russia. Forecasting based on them is inoperative on almost the whole of the non-chernozem belt, but to make up for that they are unconditionally significant for almost all the producing belt, except for the extreme South (at the moment I can offer no solution for the interesting question why there is such a

differentiation of the whole territory of former European Russia into two parts. It may either result from general failure of our indicators to forecast the weather in the ineffective regions or may instead be a consequence of a different dependence of yield on weather there from that which obtains in the producing belt. Apparently the second hypothesis should be given preference. Possibly both these conditions may be operative.)

Taking into consideration the present (This paper was written in 1940. The Editor.) division into republics and oblasts, these indicators are found to be effective: for the North of the Ukraine (Vinnitsa, Kamenets-Podolsk, Zhitomir, Kiev, Poltava, Kharkov, Donetsk (oblasts), Rostov oblast, for the Central Chernozem oblasts (Kursk, Orlov, Voronezh, Tambov, Tula, Ryazan, Penza), for the Volga oblasts (Stalingrad, Saratov, Kuibyshev, Volga German ASSR), and for the northeast region, consisting of the Bashkir, Tatar, Udmurt, Chuvash, Mari and Mordov ASSR and Chkalov, Chelyabinsk, Sverdlovsk, Kirov, Gor'kiy and Ivanovo oblasts.

Krasnodar oblast, Ordzhonikidze Kray, and Dagestan with the adjoining AO, and likewise Kazakhstan, Siberia, Central Asia and Transcaucasia were not included in the investigation (no long-term crop data).

If we follow up the degree of parallelism between the average crop yield for each year, not for all 50 guberniyas of former European Russia but only for the territory over which the indicators are effective, the parallelism is increased (but then this was to be expected). The regression equation in this case assumes the following form:

$$1Y = 24,7961 - 11,2973 X_0 + 2,0917 c; R = 0,760$$

The data on the crop yields for these 23 guberniyas, the indices $1Y$, calculated from the equation, and the divergences between them, are presented in the following table:

[See table on following page]

TABLE 12

Year	y (Observations)	1^y	$y - 1^y$	Year	y (Observations)	1^y	$y - 1^y$	Year	y (Observations)	1^y	$y - 1^y$
1883	+1.2	+9.5	-8.3	1894	+27.7	+5.2	+22.5	1905	-9.1	+16.2	-25.3
1884	+13.1	-3.7	+16.8	1895	+12.2	+14.7	-2.5	1906	-33.9	-17.5	-16.4
1885	-6.4	-2.3	-4.1	1896	+9.1	+8.2	+0.9	1907	-10.3	-19.6	+9.3
1886	-0.6	+7.8	-9.4	1897	-21.9	-11.9	-10.0	1908	-10.0	-1.3	-8.7
1887	+17.7	+26.1	-8.4	1898	-8.4	-13.1	+4.7	1909	+22.0	-4.2	+17.8
1888	+7.8	+9.2	-1.4	1899	+21.8	-11.8	+10.0	1910	+4.9	-8.7	+13.6
1889	-20.4	-16.6	-3.8	1900	+0.0	+1.1	+6.9	1911	-27.6	-19.4	-8.2
1890	-11.7	-21.8	+10.1	1901	-23.0	-14.0	-9.0	1912	+8.4	-1.4	+7.0
1891	-38.0	-27.2	-10.8	1902	+21.0	+15.3	+5.7	1913	+16.2	+1.7	+14.5
1892	-16.5	-6.7	-9.8	1903	+2.9	+21.4	-18.5	1914	-11.5	-1.3	-10.2
1893	+19.0	+6.1	+12.9	1904	+28.5	+17.9	+10.6	1915	+8.1	+7.6	+0.5

- Side -

CHAPTER II

9. As has been shown in the last chapter, the dynamics of the temporary process from autumn through winter to spring serves as a signal of the forthcoming crop yield. A steep slope of the curve that expresses the basic temperature process is an unfavorable sign, while on the contrary a moderate slope of the curve is one of the favorable signs (From all that has been said above, this characterization is accurate in the relative, correlational sense, i. e. a slanting course of the curve favors a good yield, but does not condition it as a necessary and obligatory consequence).

The former type of curve evidences an increased difference between the winter temperatures, on the one hand, and those of autumn and spring, on the other, while the latter type indicates a diminished difference. The higher the temperature in autumn and spring, and the lower it is in winter, the more sharply is the former type of curve expressed, and, conversely, the latter type corresponds to a cold autumn and spring and a warm winter.

Such are the conclusions from the analysis of the character of the temperature curves. Let us see whether they are really justified, and in what sectors of the autumn, spring and winter they fall.

We correlate the ten-day temperatures from October to April inclusive, at Ul'yanovsk station, with the indices of the forthcoming harvest. As a result we obtain the following numerical values for the correlation coefficients.

[See Table on following page]

TABLE 13

Periods		Correlation Coefficient Between Crop Yield and Mean Diurnal Temperature
October	1st 10 days	-0.146
	2nd 10 days	-0.267
	3rd 10 days	-0.311
November	1st 10 days	+0.081
	2nd 10 days	+0.207
	3rd 10 days	+0.186
December	1st 10 days	-0.042
	2nd 10 days	+0.020
	3rd 10 days	+0.059
January	1st 10 days	+0.018
	2nd 10 days	+0.193
	3rd 10 days	+0.328
February	1st 10 days	-0.159
	2nd 10 days	+0.136
	3rd 10 days	+0.019
March	1st 10 days	-0.078
	2nd 10 days	-0.100
	3rd 10 days	-0.179
April	1st 10 days	-0.064
	2nd 10 days	-0.034
	3rd 10 days	-0.392

The general character of these coefficients corresponds to the type of curve: the minus sign before a coefficient, indicating signals unfavorable for the crop yield (the warmer the weather, the worse for the yield), during the 10-day period of autumn and spring. On the contrary, the plus signs (the warmer the better) apply to the winter months from November to February inclusive: of 12 decades only two have a minus sign. Of the coefficients presented, however, the two extremes (the first decade of October and the third of April and the median (the third decade of January) attract our attention first of all.

The minus signs before the correlation coefficients for the extreme decades indicate that the higher the temperatures in these decades the less satisfactory is the crop to be awaited. For the median winter decades, on the contrary, the plus sign before the correlation coefficient speaks in favor of a direct parallelism between the temperatures and the forthcoming crop yield (the warmer the better, the colder the worse). Each of these signals, however, is only slightly effective, since the correlation coefficients are small (0.346 to 0.392). Let us see how the forecasting significance is increased by the combined use of these indicators. We add up the temperatures of the three decades selected. Since the third decade of January corresponds to a correlation coefficient with the opposite sign (a plus sign, while the other two decades, the first in October and the third in April, have minus signs), we change its sign before adding it. For instance, for 1892/93 we have the following temperatures: first decade of October, plus 2.8 degrees; third decade of April plus 3.5 degrees; third decade of January, minus 19.2 degrees. The resulting index of joint signalization is

equal to 25.5 degrees. For 1905/06 the corresponding temperatures -- plus 12.7, plus 15.9 and plus 11.1 [an obvious mispring for -11.1.] taken together give 39.7 degrees. Such indices were calculated for each year of the 33-year period. It is easy to convince oneself that the temperature difference between the autumn and spring decades, on the one hand, and the median winter decade on the other, (For instance: 1892/93: plus 2.8 plus 3.5 -(-19.2) = 25.5
1905/06: plus 12.7, plus 15.9 -(-11.1) = 39.7

We thus commence the analysis and detail study of the character of the march of the basic temperature process from autumn through winter to spring at those parts of it which serve to signal the forthcoming crop yield. The correlation coefficient between this total index and the crop yield is equal to -0.589.

But we have seen from the first chapter that the character of the basic temperature dynamics is an indicator for more than the current year, and serves for two years. Let us verify this with respect to the decades selected. We correlate their temperatures with the crop yield, but move these decades back by one agricultural year, i.e. we compare the temperatures of the first October decade in 1881 with the 1883 crop yield, the temperatures of the same decade in 1882 with the 1884 crop yield, and so on down to 1913, the temperatures in the first October decade of which are compared with the 1915 crop yield. The temperatures of the third January decade correspond to the 1883 crop yield, that decade in 1883 to the 1884 crop yield, etc. The third April decade is compared in the same way as the January decade. The coefficients so calculated have the values

$$r_{y, O-1} = -0.403 \text{ and } r_{y, A-3} = -0.254.$$

The sum of the temperatures of these two decades correlates with the crop yield to the extent of $+0.437$. In a word, in complete agreement with the conclusions of the first chapter, both extreme decades of the period from October to April inclusive, one in autumn and the other in spring, serve as signals of the crop yields in the following agricultural year. The higher their temperatures, the more favorable are their prognostic indications.

The forecasting significance of the temperature of the third January decade for the forthcoming crop yield reduces down to zero, since the coefficient r is equal to plus 0.002 . For this reason this temperature for the preceding agricultural year is excluded from the further course of the investigation. Our decipherment of the parabolic equations of Chapter I -- which reproduce in each year the peculiarities in the march of the basic temperature process from autumn through winter to spring -- has thus yielded 5 indicators.

1. Temperature of the 1st October decade of the current agricultural year -0.346
2. Temperature of the 1st October decade of the prior agricultural year -0.403
3. Temperature of the 3rd January decade of the current agricultural year -0.328
4. Temperature of the 3rd April decade of the current agricultural year $+0.392$
5. Temperature of the 3rd April decade of the prior agricultural year -0.254

These coefficients are small and fluctuate over a rather narrow range (from 0.254 to 0.403), and the individual prognostic significance of each is insignificant. By summation of these indices we form a single integral index for all five decades. These index cumulates the weak signals of its separate component elements and gives as a result a considerably greater effect. We shall designate it by the symbol T_5 . Obviously T_5 is equal to (the temperature of the third January decade is taken with the opposite sign, since its coefficient or correlation with the crop yield has a sign opposite to those of the four decades). In practice this reduces down to the simple addition of the absolute values for the temperatures of all five decades, for the minus sign before the January temperatures is turned into plus by the subtraction. The coefficient of correlation between the crop yield and T_5 is equal to -0.698.

Table 14 presents the values of O_1 , O_{-1} , J_3 , A_3 , A_{-3} , and T_5 for Ul'yanovsk meteorological station.

[See Table 14 on following page]

This equation is distinguished from the first one, given in the last chapter, by the replacement of X_0 by the index T_5 . We recall that X_0 characterizes the degree of steepness or gentleness of the march of the basic temperature process from October to April inclusive, while T_5 is an index obtained by the detailed analysis and precise study of this process during its separate intervals. The new equation definitely gives better results: the coefficient of multiple correlation is 0.776, and after elimination of the systematic error is 0.765, while its corresponding values for the first equation are

TABLE 14

Agricultural

Year	O ₁	O ₋₁	J ₃	A ₃	A ₋₃	T ₃	T ₅	y	a _{PW}	T ₆
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1882/83	4.3	3.7	-14.9	6.6	4.2	25.8	33.7	+0.5	16.2	49.9
1883/84	7.9	4.3	-10.1	3.4	6.6	21.4	32.3	+7.0	15.0	47.3
1884/85	8.2	7.9	-8.7	2.8	3.4	19.7	31.0	-10.8	13.3	44.3
1885/86	5.8	8.2	-18.9	6.1	2.8	30.8	41.8	0.0	15.8	57.6
1886/87	2.3	5.8	-12.5	4.8	6.1	19.6	31.5	+14.6	13.6	45.1
1887/88	5.3	2.3	-14.5	9.2	4.8	29.0	36.1	+9.1	14.5	50.6
1888/89	11.5	5.3	-15.5	9.1	9.2	36.1	50.6	-18.2	30.9	81.5
1889/90	9.0	11.5	-12.8	10.0	9.1	31.8	52.4	-7.3	22.4	74.8
1890/91	6.5	9.0	-26.6	8.3	10.0	41.4	60.4	-29.0	30.4	90.8
1891/92	7.7	6.5	-27.2	8.3	8.3	43.2	58.0	-13.2	21.2	79.2
1892/93	2.8	17.7	-19.2	3.5	8.3	25.5	41.5	+21.8	12.0	53.5
1893/94	11.8	2.8	-5.9	5.5	3.5	23.2	29.5	+22.0	11.2	40.7

TABLE 14 (cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1894/95	1.6	11.8	-5.5	4.1	5.5	11.2	28.5	+9.8	9.9	38.4
1895/96	10.8	1.6	-17.4	5.0	4.1	33.2	38.9	+4.8	18.4	57.3
1896/97	11.0	10.8	-14.2	6.5	5.0	31.7	47.5	-14.3	22.5	70.0
1897/98	3.6	11.0	-12.4	4.2	6.5	20.2	37.7	-1.2	15.9	53.6
1898/99	2.2	3.6	-9.3	7.4	4.2	18.9	26.7	+10.0	18.7	45.4
1899/1900	12.9	2.2	-19.0	3.6	7.4	35.5	45.1	+2.5	15.2	60.3
1900/01	9.8	12.9	-4.3	13.9	3.6	28.0	44.5	-13.9	18.6	63.1
1901/02	4.5	9.8	-11.3	4.0	13.9	19.8	43.5	+15.3	9.6	53.1
1902/03	-0.5	4.5	-15.1	12.9	4.0	27.5	36.0	+4.0	12.5	48.5
1903/04	3.6	-0.5	-7.5	11.7	12.4	22.8	35.2	+18.9	8.1	43.3
1904/05	7.8	3.6	-13.3	11.5	11.7	32.6	47.9	-7.4	13.7	61.6
1905/06	12.7	7.8	-11.1	15.9	11.5	39.7	59.0	-23.9	21.4	80.4
1906/07	2.4	12.7	-15.0	11.0	15.9	28.4	57.0	-10.0	17.0	74.0
1907/08	6.2	2.4	-9.4	9.1	11.0	24.7	38.1	-8.5	27.2	65.3
1908/09	4.6	6.2	-12.1	6.3	9.1	23.0	38.3	+15.7	12.9	51.2

TABLE 14 (cont'd)

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
1909/10	10.0	4.6	-10.5	13.5	6.3	34.0	44.9	+6.4	14.1	59.0
1910/11	4.5	10.0	-15.6	10.2	13.5	30.3	53.8	-18.2	15.4	69.2
1911/12	6.1	4.5	-14.0	3.8	10.2	23.9	38.6	+7.9	12.7	51.3
1912/13	4.8	6.1	-13.0	9.4	3.8	27.2	37.1	+17.1	18.0	55.0
1913/14	3.6	4.8	-9.3	7.4	9.4	20.3	34.5	-11.5	5.7	40.2
1915/15	5.9	3.6	-4.4	7.5	7.4	17.8	28.8	+10.9	21.2	50.0
r =	-0.346	-0.403	+0.328	-0.392	-0.254	-0.589	-0/698	--	-0.596	-0.754

Symbols used in Table 14:

O_1 : mean diurnal temperatures of first October decade (meaning the October of the first figure in the current agricultural year. E.e. the line 1882/83 contains, in this column, the mean diurnal temperature of the 1st October decade of 1882);

O_{-1} : temperature of first October decade of the preceding agricultural year;

J_3 : temperature of the third January decade (in the year shown by the second figure of the current agricultural year, e.g. on the line 1882/83, the temperature of the third decade of January, 1883);

TABLE 14 (cont'd)

A_3 : temperature of the third April decade (see note on J_3);

A_{-3} : temperature of the third April decade of the preceding agricultural year;

$$T_3 = O_1 - J_3 + A_3;$$

$$T_5 = O_1 + O_{-1} - J_3 + A_3 + A_{-3};$$

y : yield in percent departure from norms

a^P_W : temperature difference between the last two autumn decades and the first two winter decades;

$$T_6 = O_1 + O_{-1} + J_3 + A_3 + A_{-3} + a^P_W.$$

0.732 and 0.704 respectively. Moreover, a lesser weight is assigned in it (the new equation) to the prior agricultural year than to the current agricultural year of the forthcoming harvest, for the first October decade and the third April decade participate in both years, while the third January decade is omitted for the preceding year: it has already been noted that the correlation coefficient between y and the sum of O_{-1} and A_{-3} is equal to -0.437, while for the algebraic sum $O_1 - J_3 + A_3$ it attains the value of -0.589. The former year thus gives a weaker signal than the current agricultural year.

We are thus left with the five indices mentioned, which make up the integral index T_5 . The inclusion of the remaining decades in the investigation would be very venturesome, in view of the negligible values of the correlation coefficients for the November, December, February and March decades, and the first two in April. It would involve the more danger since neither regular growth nor decline is observed for them and even their signs do not always completely agree (the series of decades with plus signs is twice interrupted by "minus decades"). (It was unfortunately impossible for the author to carry out the same detailed work for the decades of the "past" agricultural years as for those of the agricultural year immediately preceding the harvest, and he therefore confined himself to the mere verification of the forecasting significance of only 3 decades (the 1st October, 3rd January and 3rd April)).

The two October decades with increased coefficients, respectively -0.267 and -0.311, still remain. These are the second and third decades in this month, and they are close both in value and in sign to the first October decade. They have not been included in this study because they form a component element every year in the

characterization of a phenomenon whose significance will now be elucidated.

10. When he analyzed the conditions that determine the degree of steepness in the march of the underlying temperature process from November to the end of April, the author posed the question whether the peculiarities of the transition from autumn to winter to spring do not constitute one of the elements that influence the temperature dynamics represented by the paraboles. The transition from winter to spring did not yield sufficiently definite and clear results, but for the autumn shift in the weather the results proved to be positive. The change from autumn to winter may be abrupt or gradual. The temperature difference between the last two warm decades and the first two cold ones is adopted as an index of this sharpness. Thus, for instance, the temperature in the last two warm decades of the autumn of 1890 (thus in the agricultural year 1890/91) -- which were the 3rd in October and the 1st in November -- were plus 2.5 and plus 2.3 degrees respectively, while the first two winter decades that followed -- the 2nd and 3rd in November -- had temperatures of -8.0 and -17.6 degrees respectively. The temperature difference, which we denote by a_{Pw} , was equal to $+ 2.5 + 2.3 - (-8.0) - (-17.6) = 30.4$ (All of this obviously amounts to adding up the temperatures without distinction of signs.)

For the autumn of 1900 (and consequently the agricultural year 1900/01) the last two October decades had temperature of plus 6.1 and plus 2.6 respectively and the first two November decades had -2.8 and -7.1: as a result a_{Pw} equalled $+ 6.1 + 2.6 - (-2.8) - (-7.1) = 18.6$. Such indices were calculated for each year of the period 1883-1915.

On correlating aP_w with the crop yield, we obtain a correlation coefficient that testifies by its magnitude to an undoubted parallelism between the values of aP_w and the forthcoming harvest. This coefficient is equal to -0.589 . Its minus sign shows that a sharp transition from autumn to winter is a bad sign, and, on the other hand, that a gentle change of autumn into winter is a favorable sign.

Inclusion of this criterion in the integral index T_6 increases the coefficient of correlation to -0.754 . We note that the index T_6 constructed by the same method for the meteorological station at Kaxan' gives a coefficient of correlation of -0.729 with the crop yield. The regression equation connecting y (the yield) with T_6 and B_c (the average date of unfolding of the leaves on the trees) is of the following form:

$$y = 10,6492 - 0.7625 T_6 + 1,0529 B_c ; R = 0.824 \quad (5)$$

The coefficient of multiple correlation, corrected for systematic error, is 0.811 for this equation and is thus appreciably higher than the same coefficient (0.704) corresponding to the regression equation:

$$y = 13,6364 - 8.1135 X_c - 1.6525 B_c \quad (\text{See Chapter I})$$

It is clear that the replacement of the index X_c by the index T_6 leads to an improvement in the forecasting system. The latter is distinguished from the former, as results from all that has been set forth above, by the more detailed determination,

throughout the entire course of the 7-month period, of the time intervals that are prognostic significance for the forthcoming harvest. These intervals are the following: 1st October decade, 3rd January decade, 3rd April decade, and the transition from autumn to winter (4 decades); together with the October decade and the 3rd April decade of the autumn/winter/spring of the prior agricultural year. In contrast to the index T_6 , the index X_6 gives only a general characterization, based on the temperatures of all 21 decades, of the temperature dynamics from autumn through winter to spring, and does not segregate the "active" decades from those that are neutral in their prognostic effect.

CHAPTER III

11. We shall now dwell on the active decades and examine their climatological significance more in detail in the region of the lower Kama, and the Middle Volga (Kazan' Station and Ul'yanovsk). The thirty-year period from 1882 to 1915 permits us to establish the following state of affairs. The first October decade is the last decade in this territory that is always autumnal. In fact, all of the 34 first October decades at Kazan Station are characterized by a mean diurnal temperature above zero, while at Ul'yanovsk there was only one slight exception--the autumn of 1902, with a temperature of -0.5 for the first decade of October. The second decade, and even more so the third decade, is not always autumnal; in 1882, 1898, 1903, 1908, and 1911, the second decade of October had a temperature under zero, while this was true at Ul'yanovsk Station in 1882, 1898, 1908, 1910 and 1911. At both stations, a temperature below zero is noted for the third decade of October in 16 cases out of the 34.

The other extreme decade of the period under study--namely the third April decade, is always a spring decade under the climatic condition: In all the years studied, it was characterized by a temperature above zero at both stations. The second decade of April is not always a spring decade; in 6 cases out of the 34, Kazan Station observed mean diurnal temperatures below zero for this decade, while at Ul'yanovsk 5 such years were noted. Finally, the first April decade has a winter temperature in 15 of the years at Ul'yanovsk and 13 of the years at Kazan, and a spring temperature in 19 and 21 cases respectively. A second characteristic feature of the extreme decade during the period under study is the stability of their temperature conditions with respect to the two other decades of the same month. In

26 cases, the highest mean diurnal temperature throughout the decade, out of all three October decades, falls in the first decade in October, and in the 8 remaining cases, it falls in the second decade at Uli'yanovsk Station, while the corresponding figures for Kazan were 27 and 7 respectively. The third October decade does not show a single maximum. The situation of the third April decade relative to the temperatures of the other two April decades is even more definite: In 31 cases the highest decade temperature falls in the third decade of April, and on the second decade in 3 cases, while it was not observed to fall on the first decade in even a single year. The last "active decade" is the third January decade, which is one of the "central" winter decades.

We shall now modify some of the bases of selection of the "active" decades, as follows: Instead of a fixed calendar decade, the first in October, we shall now take in each year the October decade with the highest temperature. Instead of the third April decade we shall include the warmest April decade in our study and, finally, instead of the third calendar decade in January, we shall use the coldest of the three decades which include the second and third January decades and the first February decade. In practice this slightly modifies the series O_1 and A_3 , for in the vast majority of cases, the first October and third April decades still go into this series, and only rarely do the second decades go there. As for the third of October decade and the first April decade, the probability of their being warmer than the other decades in the month is exceedingly small. The series G under those the most modification of all for the minimum decade temperatures in general are distributed very much at random throughout of the course of the winter and especially during the period from the second January decade to the first February decade inclusive. But this modification in the basis of selection is of a certain importance for

the forecasting, since it rejects the relatively narrow decade basis (which can be distorted in this year or in that by every kind of accidental circumstance), and in essence means a transition to a monthly basis, since the selection of a decade is made on the basis of a comparison of the temperatures of all three decades. At the same time, this method does not desolve in the mean monthly indices those relatively short outbreaks of warmth or cold which may be of significance for prognosis. The index P_{a-w} (transition from autumn to winter), which is based on four decades, is no longer needed in the expanded basis. We also give up our observations at only a single station and pass to the use of an index calculated for two stations, Ul'yanovsk and Kazan. At the end of the second chapter, we showed that T_6 , set up on data of Kazan, yields the same result obtained for Ul'yanovsk. This result is not surprising, since the parallelism of temperatures between these stations is very great. However, since we are speaking of prognostic indices that manifest themselves of the territories of the lower Kama and the Middle Volga all of the material at the author's disposition relating to this region, and worked up by him, should be utilized. Kazan characterizes the lower Kama, Ul'yanovsk characterizes the Middle Volga, and they jointly characterize the region of the lower, and the Middle Volga. The soundness of the conclusions only increases when the observations of two stations are used instead of one. Table 15 gives indices calculated for both Ul'yanovsk and Kazan Stations.

In this table all of the previous indices are presented, worked out in the variant indicated in this chapter: For the two stations Ul'yanovsk and Kazan together. Besides this the new prognostic signs are introduced, concerned with the differences in rain-

fall for November and January and April for the current agricultural year between the east and southwest. The significance of these signs will be discussed later. As may be seen from this table the correlation coefficients of the individual indices under O , O_{-1} , A , A_{-1} , J , etc, and hardly different from the values of the corresponding correlation coefficients for the first variant, worked out only for Uliyanovsk Station and thus on a narrower base. (The values O_1 , O_{-1} , A , A_{-1} , and J , belong to entirely definite decades, while the second variant dispenses with any attachment to specific calendar decades.)

In exactly the same way, the coefficients of multiple correlation of the new regression equation (6) coincide with the coefficient of the multiple correlation of equation (5);

$$r = 22.5291 - 0.8291T_2 + 0.9110B_2 + 0.829 \quad (a)$$

Elimination of the systematic errors lowers the coefficient to the level of the corrected correlation coefficient of the analogous regression equation of the first variant (5), namely: 0.816 and 0.811 respectively. Thus the more precise variant entirely reinforces the conclusions of the last chapter. (It would consequently be completely futile to try to "wave aside" the results obtained, by referring to the fact that the whole study was founded only on the data of a single station and based only on a limited number of calendar decades. The Kazan' and Uliyanovsk stations, which are far enough away from each other, jointly characterize the temperature regime of an area that is by no means small.)

12. In this paper we have already established that the temperatures of the first October decade and the last April decade at Uliyanovsk Station constitute symptoms that serve to signal the

coming crop yields. The investigation has likewise shown that these same decades of the preceding year also possess prognostic significance. We present the coefficients of correlation with the crop yields of the temperatures during these decades at Ul'yanovsk Station:

$$y, c_1 = -0.346 \text{ (cf p. 72)}; y, A_3 = -0.372,$$

$$y, c_2 = -0.403,$$

$$y, A_3 = -0.254$$

Further investigation showed that analogous indices were also obtained for Kazan' Station. All of this permits us to draw the conclusion that these signs, manifesting themselves over the territories of the lower reaches of the Kama and the middle courses of the Volga, possess prognostic significance. The coefficient of correlation has a minus sign everywhere, which testifies to the fact that high temperatures are a sign unfavorable to them. Each of these signs is only a feeble prognostic, but use of their totality increases their prognostic significance. Thus, for example, if the crop yield be correlated with the sum of the temperatures for the first October decade and the last April decade, the correlation coefficient rises to -0.508 , while the sum of the temperatures for all four decades already correlates with the crop yield to the extent of -0.648 .

The question arises whether the prognostic effect of these indices is confined to the region noted (lower Kama and middle Volga) or whether this region should be expanded.

With this in view we correlated the crop yield with the temperature of the first October decade and with the temperature of the third April decade for two years at a series of stations (for which

the data was available to me) and obtained the following numerical values for these coefficients:

[Table 16 on the following page]

All these stations are located on the territory of the so-called producing belt of the European part of the USSR. (North of this region I have data only for the stations at Pinsk, Leningrad, Arkhangel'sk and Moscow, and therefore decided not to undertake the going beyond the limits of the chernozem belt.)

[Table 15 on second following page]

The following conclusions flow from these tables of correlation coefficients: (a) for all 18 stations and all 4 decades the correlation coefficients have the minus sign, thus confirming the conclusion drawn by us for Ul'yanovsk and Kazan', namely: high temperatures in the first October decade and the third October decade for the two years preceding the harvest constitute an unfavorable sign for the coming crop yield, while low temperatures, on the contrary, are a favorable sign. The numerical values of the correlation coefficients are small, just as they were for Ul'yanovsk and Kazan', which is a proof of the low prognostic value of each index taken separately; (b) an entirely definite regional effectiveness of the indices is noted, and varies for each of them. The East is outstanding in effectiveness for the first October decade: the four easternmost stations (Sverdlovsk, Zlatoust, Polibino and M. Uzen') have the highest correlation coefficients, ranging from -0.402 to -0.455, which are not repeated at the other stations lying to the west (see Base Map 2). Six stations with indices from -0.328 to -0.386 (Molotovsk, Kirov, Kazan', Ul'yanovsk,

TABLE 16

	$r_{y,0}$	$r_{y,0-1}$	r_{y,A_3}	$r_{y,A-3}$
Molotovsk (Perm')	-0.362	-0.397	-0.222	-0.220
Sverdlovsk	-0.402	-0.417	-0.167	-0.191
Zlatoust	-0.428	-0.414	-0.151	-0.184
Kirov (Vyatka)	-0.328	-0.363	-0.252	-0.305
Kazan'	-0.329	-0.392	-0.409	-0.265
Polibino	-0.455	-0.367	-0.365	-0.257
Ul'yanovsk (Simbirsk)	-0.346	-0.403	-0.392	-0.254
Nikolayevskoye (Saratov oblast)	-0.386	-0.426	-0.372	-0.249
M. Uzen'	-0.428	-0.428	-0.385	-0.236
Zemetchino	-0.305	-0.415	-0.441	-0.215
Tambov	-0.261	-0.427	-0.438	-0.238
Voronezh	-0.245	-0.400	-0.469	-0.177
Uryupino	-0.371	-0.384	-0.398	-0.237
Kharkov	-0.301	-0.392	-0.392	-0.253
Lugansk	-0.305	-0.413	-0.416	-0.307
Kirovo (Yel'izavetgrad)	-0.236	-0.226	-0.392	-0.210
Kiev	-0.167	-0.101	-0.340	-0.191
Odessa	-0.127	-0.120	-0.286	-0.239

TABLE 15

Agricultural

Year	O	O ₋₁	A	A ₋₁	J	O _c	N
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1882/83	4.2	8.0	6.2	4.4	-23.6	10.7	-5.3
1883/84	7.8	4.2	3.0	6.2	-17.1	5.1	-8.0
1884/85	8.5	7.8	3.3	3.0	-16.5	4.4	-9.6
1885/86	6.0	8.5	6.0	3.3	-19.0	5.5	-9.7
1886/87	3.2	6.0	5.5	6.0	-15.3	5.8	-8.9
1887/88	4.8	3.2	9.4	5.5	-16.2	2.4	-12.7
1888/89	10.7	4.8	9.3	9.4	-16.7	18.0	-13.2
1889/90	10.2	10.7	10.3	9.3	-13.6	19.9	-5.1
1890/91	6.2	10.2	8.2	10.3	-26.0	5.1	-28.8
1891/92	7.7	6.2	8.0	8.2	-27.3	11.7	-10.0
1892/93	5.3	7.7	3.5	8.0	-19.7	6.5	-3.9
1893/94	11.7	5.3	5.0	3.5	-14.2	3.0	-8.5
1894/95	4.4	11.7	3.8	5.0	-14.7	6.0	-4.6
1895/96	10.7	4.4	5.8	3.8	-20.7	11.7	-5.7
1896/97	10.7	10.7	5.8	5.8	-15.8	14.5	-9.4

(Page 2 Table 15)

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1897/98	7.5	10.7	4.1	5.8	-12.2	11.2	4.6
1898/99	2.2	7.5	7.6	4.1	-9.9	13.4	-5.0
1899/1900	12.5	2.2	3.5	7.6	-19.8	3.1	-12.9
1900/01	9.3	12.5	13.2	3.5	-15.6	8.5	-10.1
1901/02	5.0	9.3	2.9	13.2	-17.3	2.4	-8.2
1902/03	1.0	5.0	12.9	2.9	-15.4	4.6	-7.6
1903/04	2.8	1.0	11.3	12.9	-14.2	2.2	-4.9
1904/05	7.5	2.8	11.3	11.3	-13.9	7.5	-6.9
1905/06	11.8	7.5	16.1	11.3	-11.6	15.0	-6.4
1906/07	6.4	11.8	11.0	16.1	-20.5	8.0	-7.8
1907/08	6.4	6.4	8.9	11.0	-13.8	7.4	-19.5
1908/09	4.7	6.4	6.4	8.9	-12.5	11.5	11.5
1909/10	9.8	4.7	13.4	6.4	-13.2	5.6	-6.8
1910/11	4.7	9.8	9.7	13.4	-22.0	9.1	-2.5
1911/12	6.0	4.7	3.6	9.7	-16.7	7.1	-4.9
1912/13	4.6	6.0	9.6	3.6	-14.8	8.7	-3.8
1913/14	3.3	4.6	7.1	9.6	-15.6	4.5	-1.2
1914/15	5.8	3.3	7.6	7.1	-10.0	8.8	-11.6
r =	-0.344	-0.379	-0.385	-0.286	+0.334	-0.404	+0.372
n =	2,986	2,988	3,438	3,484	4,078	4,429	5,163

(page 3 of Table 15)

a^p_w	T_6	B	B_{-1}	B_c	P_{west}	P_{east}	$west\ P_{east} = Y$ $=(P_{east}-P_{west})$
[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16] [17]
16.0	62.4	30	31	30.5	146.6	114.9	-31.7 +0.5
13.1	51.4	27	30	28.5	76.4	95.8	+19.4 +7.0
14.0	53.1	26	27	26.5	105.0	87.5	-17.5 -10.8
15.2	58.0	42	26	34.0	81.6	42.0	-39.6 00.0
14.7	50.7	37	42	39.5	70.3	65.0	-5.3 +14.6
15.1	54.2	28	37	32.5	126.4	82.4	-44.0 +9.1
31.2	82.1	34	28	31.0	140.7	77.7	-63.0 -18.2
24.9	79.0	23	34	28.5	75.8	80.6	+4.8 -7.3
33.9	94.8	31	23	27.0	162.9	39.0	-123.9 -29.0
21.7	79.1	35	31	33.0	95.1	93.6	-1.5 -13.2
10.4	54.6	40	35	37.5	106.7	80.3	-26.4 +21.8
1155	51.2	32	40	36.0	76.2	65.9	-10.3 +22.0
10.6	50.2	34	32	33.0	116.5	75.6	-36.9 +9.8
17.4	62.8	42	34	38.0	68.5	44.4	-24.1 +4.8
23.9	72.7	32	42	37.0	111.8	86.4	-25.4 -14.3

(Page 4 of Table 15)

[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]
15.8	56.1	29	32	30.5	67.9	56.1	-11.8	-1.2
18.4	49.7	32	29	30.5	30.0	80.3	+50.3	+10.0
16.0	61.6	35	32	33.5	89.6	80.8	-8.8	+2.5
18.6	72.7	37	35	36.0	95.2	43.3	-51.9	-13.9
10.6	58.3	36	37	36.5	47.8	97.8	+50.1	+15.3
12.2	49.4	31	36	33.5	62.1	68.6	+6.5	+4.0
7.1	49.3	38	31	34.5	80.4	77.0	-3.4	+18.9
14.4	61.2	37	38	37.5	127.1	64.7	-62.4	-7.4
21.4	79.7	23	37	30.0	158.2	51.9	-106.3	-23.9
15.8	81.6	35	23	29.0	183.2	78.9	-24.3	-10.0
26.9	73.4	34	35	34.5	129.0	36.1	-92.9	-8.5
13.0	51.9	36	34	35.0	66.0	102.9	+36.9	+15.7
12.4	59.9	27	36	31.5	127.7	122.7	-5.0	+6.4
11.6	71.2	29	27	28.0	106.9	77.0	-29.9	-18.2
12.0	52.7	35	29	32.0	119.8	92.2	-27.6	+7.9
17.5	56.1	25	35	30.0	76.8	112.3	+35.5	+17.1
5.7	45.9	17	25	21.0	118.8	92.9	-25.9	-11.5
20.6	54.4	26	17	21.5	150.6	105.6	-45.0	+10.9
-0.592	-0.777	+0.263	+0.267	+0.3	-1.517	+0.387	+0.635	-
6.260	-	5.649	5.554	55	32.38	22.13	39.213	-

(Page 5 of Table 15)

Symbols used in Table 15:

O: temperature of warmest October decade of current agricultural year;

O₋₁: temperature of warmest October decade of prior agricultural year;

A: temperature of warmest April decade of current agricultural year;

A₋₁: temperature of warmest April decade of prior agricultural year;

J: temperature of the coldest decade of the three : 2nd and 3rd January decades and 1st February decade;

O_c: Sum of mean temperatures of last two autumn decades;

N: Sum of mean temperatures of first two winter decades;

aPw: (O_c-N) : transition from autumn to winterT₆: O O J A A P

B: Time of unfolding of leaves on trees at B. Fenino Station;

B₋₁: Same for prior year;B_c 0.5 (B B) ;P_{west}: total precipitation for November, January and April at Odessa and Kiev Stations (southwest)P_{east}: " " " " " " " " " Kazan' and Ul'yanovsk stations (east)east^Pwest: difference between the rainfall for these months (east minus west) i. e.

$$\text{east}^P\text{west} = P_{\text{east}} - P_{\text{west}}$$

Nikolayevsk, Uryupino) immediately adjoin this region. On Base Map 2 the first region is bounded on the west by the double line ; the second region is located between the single line and the double line . From the eight remaining stations, with correlation coefficients from -0.127 to -0.305, the region of the Southwest, with the lowest coefficients -- -0.127 for Odessa and -0.167 for Kiev -- is separated by the broken line . Thus the prognostic value of the temperature in the first October decade is very definitely increases from the Southwest of the European part of the USSR towards the east. Owing to my lack of material, I am unfortunately unable to say how far east this rise in prognostic value of the first October decade temperature continues to prevail.

[See Base Map 2 on following page]

[See Base Map 3 on second following page]

Let us now examine the regional distribution of the second index: the temperature in the third April decade. We turn to Base Map 3. The central position of the region of highest effectiveness stands out very clearly. Within the limits of this region, bounded by the single lines , the correlation coefficients range from -0.365 to -0.469. The Northeast is characterized by coefficients that are definitely low -- -0.252 for Kirov, -0.222 for Molotovsk, -0.167 for Sverdlovsk, -0.151 for Zlatoust. Values as low as these are not found anywhere to the southwest of this region. The two westernmost stations of Odessa and Kiev likewise have lowered correlation coefficients (-0.286 and -0.340) in comparison to the central region (though this reduction is not so considerable as in the northeast). The lack of data for stations to the west of

Odessa and Kiev prevents our following the subsequent course of this lowering. (It would be obviously be necessary to make use of the observation of foreign stations in this case.) In the most central region the nucleus of the whole area is clearly delineated, and has been indicated by a double line on the Base Map. The four stations it includes -- Zametchino, Tambov, Voronezh and Lugansk -- have the highest correlation coefficients, ranging from -0.416 to -0.469. Such values are not met outside the limits of this nucleus.

[See Base Map 4 on following page]

[See Base Map 5 on second following page]

We turn next to the determination of the regional distribution of the third index: the temperature of the first October decade in the agricultural year that precedes the crop (see Base Map4). Here a regional distribution is discovered in the sense of a sharp distinction of the extreme southwest, with its low correlation coefficients (Odessa, -0.120; Kirovo (Yelizavetgrad), -0.226; Kiev, -0.101) from the center and the east. The element shared in common by the first indices (i.e. the temperature in the first October decade of the same agricultural year as the crop being forecast) and this index is precisely this separation of the Southwest, in contrast to the absence of differentiation of the East and center in the latter case, and its presence for the first criterion (compare Base Maps 2 and 4).

Finally, the fourth sign -- the temperature of the fourth April decade of the year preceding the crop -- reveals no definite regional distribution. (See Base Map 5). In general, lowered cor-

Odesa and Kiev prevents our following the subsequent course of this lowering. (It would be obviously be necessary to make use of the observation of foreign stations in this case.) In the most central region the nucleus of the whole area is clearly delineated, and has been indicated by a double line on the Base Map. The four stations it includes -- Zametchino, Tambov, Voronezh and Lugansk -- have the highest correlation coefficients, ranging from -0.416 to -0.469. Such values are not met outside the limits of this nucleus.

[See Base Map 4 on following page]

[See Base Map 5 on second following page]

We turn next to the determination of the regional distribution of the third index: the temperature of the first October decade in the agricultural year that precedes the crop (see Base Map). Here a regional distribution is discovered in the sense of a sharp distinction of the extreme southwest, with its low correlation coefficients (Odessa, -0.120; Kirovo (Yelizavetgrad), -0.226; Kiev, -0.101) from the center and the east. The element shared in common by the first indices (i.e. the temperature in the first October decade of the same agricultural year as the crop being forecast) and this index is precisely this separation of the Southwest, in contrast to the absence of differentiation of the East and center in the latter case, and its presence for the first criterion (compare Base Maps 2 and 4).

Finally, the fourth sign -- the temperature of the fourth April decade of the year preceding the crop -- reveals no definite regional distribution. (See Base Map 5). In general, lowered cor-

Odesa and Kiev prevents our following the subsequent course of this lowering. (It would be obviously be necessary to make use of the observation of foreign stations in this case.) In the most central region the nucleus of the whole area is clearly delineated, and has been indicated by a double line ===== on the Base Map. The four stations it includes -- Zametchino, Tambov, Voronezh and Lugansk -- have the highest correlation coefficients, ranging from -0.416 to -0.469. Such values are not met outside the limits of this nucleus.

[See Base Map 4 on following page]

[See Base Map 5 on second following page]

We turn next to the determination of the regional distribution of the third index: the temperature of the first October decade in the agricultural year that precedes the crop (see Base Map4). Here a regional distribution is discovered in the sense of a sharp distinction of the extreme southwest, with its low correlation coefficients (Odesa, -0.120; Kirovo (Yelizavetgrad), -0.226; Kiev, -0.191) from the center and the east. The element shared in common by the first indices (i.e. the temperature in the first October decade of the same agricultural year as the crop being forecast) and this index is precisely this separation of the Southwest, in contrast to the absence of differentiation of the East and center in the latter case, and its presence for the first criterion (compare Base Maps 2 and 4).

Finally, the fourth sign -- the temperature of the fourth April decade of the year preceding the crop -- reveals no definite regional distribution. (See Base Map 5). In general, lowered cor-

relation coefficients correspond to this sign, but even these lowered coefficients are not distributed in any definite order over the territory included in this study. For instance, the highest coefficients (-0.305 and -0.307) belong to the stations at Kirov (Vyatka) and Lugansk, which are very far apart. The stations with lowered coefficients likewise disclose no definite regional distribution: Sverdlovsk, Zlatoust, Voronezh and Kiev have correlation coefficients under -0.200.

It is entirely clear from all that has been set forth that the supplementary study of 16 stations (other than Ul'yanovsk and Kazan') has expanded the basis of the analysis by allowing us to pass from an investigation based on only two stations to the use of data from a more considerable number of stations and also to a utilization of such data that is more rationally organized on a basis of regional distribution. The combination of stations forming a part of the most effective forecast regions automatically suggests itself. Such regions are the extreme East, characterized by the stations at Sverdlovsk, Zlatoust, Polibino and M. Uzen' for the first October decade and the nucleus of the central region, with the stations at Zemetchino, Tambov, Voronezh and Lugansk for the third April decade. For the first October decade we take the average temperature for each year from the reports of the former group of stations (Sverdlovsk, Zlatoust, Polibino, M. Uzen'), while for the third April decade we take the data from the reports of the second group of stations (Zemetchino, Tambov, Voronezh, Lugansk) and find the correlation coefficients of these indices with the crop yields. We obtain as a result:

$$r_{y, \bar{y}} = -0.446$$

$$r_{y, A-5} = -0.456$$

$$r_{y, A-1} = -0.425$$

$$r_{y, A-1} = -0.216$$

As we see the correlation coefficients obtained here are of the same order as those for Ul'yanovsk alone; all of them are small, which indicates that the indices selected have by themselves only weak individual forecasting power, with respect to the forthcoming crop yield. But using them as an integral whole increases their significance in this respect: thus, for instance, the sum of the temperatures of the first October decade and the third April decade correlates with the crop yield to the extent of -0.581 , while the sum of the temperatures of all four decades gives a coefficient of correlation with the yield that is equal to -0.700 , somewhat higher than the corresponding coefficient for Ul'yanovsk alone, which is -0.648 .

We now expand the territorial base: instead of the four stations for O and the four stations for A we take all the stations in the regions of effectiveness. The following 10 stations belong to the eastern region (for O): Molotovsk, Kirov (Vyatka), Sverdlovsk, Zlatoust, Polibino, Kazan', Ul'yanovsk, Nikolayevskoye (Saratov oblast), M. Uzen', and Uryupino. The central region (for A) includes the following 12 stations: Kazan', Ul'yanovsk, Polibino, Nikolayevskoye, Uryupino, Zemetchino, Tambov, Voronezh, Kharkov, Kirovo, (Yelizavetgrad), Lugansk, M. Uzen'.

As a result we now obtain the following correlation coefficients:

$$r_{y, O_1} = -0.405$$

$$r_{y, A_3} = -0.432$$

$$r_{y, O_1} = -0.424$$

$$r_{y, A_3} = -0.268$$

The sum of the temperatures for the four decades ^ccorrelates with the crop yield to the extent of -0.694. This correlation coefficient almost coincides with that for the four decades of the preceding variant (-0.700) and is only slightly different from the variant obtained for Ul'yanovsk alone (-0.648).

Thus the working up of the question with the data of 18 stations not only confirms the conclusions drawn from the work on two stations alone (Ul'yanovsk and Kazan') but also expands the basis for these conclusions and defines the place for two stations in the work of making a regional distribution. (The author was unfortunately unable, in view of the large amount of work that would have been involved, to make use of this improved variant based on regional distributions. Here again the methodological improvement of procedures leads not to deterioration but to a certain improvement of the results.)

CHAPTER IV

Let us see whether the results cannot be improved by including new indices. We turn to the rainfall throughout the 6-month period from November to April inclusive. The total precipitation during this interval at Kazan' and Ul'yanovsk stations correlates positively with the crop yield, though the correlation coefficient is not large (plus 0.331). At the same time, the total precipitation at the two southwest stations of Odessa and Kiev, during the same 6-month period, correlates in the opposite sense with the yield; the correlation coefficient here is -0.333. The conclusion suggests itself that the difference in rainfall between east and southwest has forecasting value. The more decided this difference is to the disadvantage of the East, the more disturbing is the forecast of the harvest to be expected. And, on the contrary, the more this difference is evened out, passing over infrequently even to a positive balance in favor of the East (1894, 1899, 1902, 1913, 1914), the more favorable the forecast is in general (but only in general). The correlation coefficient between the difference of the eastern and southwestern precipitation and the crop yield is plus 0.499. The regression equation establishing a parallelism between the yield and the individual components of this difference is of the following form:

(7) $y = 0.3445 + 0.0743 P_{\text{east}} - 0.0568 P_{\text{west}}; R = 0.502$

The coefficient of multiple correlation is equal to 0.502 and shows hardly any difference from the correlation coefficient for the simple difference of the precipitation (plus 0.499). There is thus no need to complicate the question by introducing two variables instead of one. This equation, however, enables us to determine the coefficients of partial correlation between y and P_{west} and P_{east} .

That of the yield with the precipitation in the East (P_{east}), eliminating the precipitation in the Southwest, is plus 0.400, while that for the precipitation in the southwest, eliminating the precipitation in the East, is -0.398. In a word, each component should be recognized, with very high probability, as an actually existing indicator-symptom of the coming crop yield, and even more does this apply to the resultant effect that represents the difference between these components.

Table 17 gives the sums of the 6-month precipitation (November, December, January, February, March, April) at Uliyanovsk and Kazan' P_E , at Odessa and Kiev P_W and the difference between them ($P_E - P_W$) designated by the symbol $E P_W$.

Thus we may include the difference in the 6-month precipitation of East and Southwest as an index-symptom. Let us attempt to detail the conclusions for the separate intervals that compose this 6-month period, as we did with respect to the temperature.

[Table 17 on following page]

It develops that forecasting value may be attributed not to the entire period as a whole, but only to three months of it: November, January and April. The correlation coefficients between the crop yield and the East-West difference in precipitation are over 0.30 for each of these three months and have plus signs, while the other three months fluctuate between -0.110 and plus 0.115:

November	plus 0.321
December	minus 0.110
January	plus 0.337
February	plus 0.115

TABLE 17

[1] Year	[2] y	[3] 6P_6	[4] 6P_3	[5] 6P_3
1882/83	+0.5	358.1	508.4	-150.3
1883/84	+7.0	288.1	311.6	-23.5
1884/85	-10.8	302.5	393.4	-90.9
1885/86	0.0	132.3	398.8	-266.5
1886/87	+14.6	288.5	365.0	-76.0
1887/88	+9.1	305.0	564.1	-259.1
1888/89	-18.2	323.3	559.1	-235.8
1889/90	-7.3	245.9	214.4	-68.5
1890/91	-29.0	197.2	448.4	-251.2
1891/92	-13.2	297.1	376.7	-79.6
1892/93	+21.8	331.9	412.6	-80.7
1893/94	+22.0	298.0	271.2	+26.8
1894/95	+9.8	328.3	508.4	-180.1
1895/96	+4.8	172.4	362.0	-189.6
1896/97	-14.3	277.8	423.9	+146.1
1897/98	-1.2	167.0	310.5	-143.5
1898/99	+10.0	269.7	222.9	+46.8
1899/1900	+2.5	232.4	466.6	-234.2
1900/01	-13.9	241.4	510.3	-268.9
1901/02	+15.3	350.0	272.5	+77.5
1902/03	+4.0	228.0	278.4	-50.4
1903/04	+18.9	258.4	324.2	-65.8
1904/05	-7.4	208.7	339.3	-130.6
1905/06	-23.9	244.7	480.3	-235.6
1906/07	-10.0	268.9	414.4	-145.5

[Table continued on following page]

Table 17 (continued)

[1]	[2]	[3]	[4]	[5]
1907/08	-8.5	252.3	481.8	-229.5
1908/09	+15.7	308.8	359.3	-50.5
1909/10	+6.4	359.6	392.9	-33.3
1910/11	-18.2	254.8	316.6	-61.8
1911/12	+7.9	322.7	442.2	-119.5
1912/13	+17.1	402.1	258.8	+143.3
1913/14	-11.5	399.0	375.3	+23.7
1914/15	+10.9	387.4	535.0	-147.6
r =		+0.331	-0.333	+0.494

March plus 0.033
 April plus 0.446

The exclusion of December, February and March is thus dictated by the results of detailed analysis of the study.

We confine ourselves to the November, January and April precipitation for East and Southwest and find the coefficient of correlation between their difference (East-West) [and the yield] to be plus 0.635. This difference is shown in the E^P_W column of Table 15.

We find the regression equation that embodies the parallelism between the yield and the three indices T , B_c and E^P_W . It has the following form:

$$Y = 19.37545 - 0.61108 T_c + 0.74630 B_c + 0.09208 E^P_W; \quad (8)$$

$$R = 0.847$$

Elimination of the systematic error lowers the coefficient of correlation to 0.830. The regression equation presented corresponds to the variant in which both temperature criteria and "precipitation" indices are shown for the separate time intervals. In the first chapter we found the regression equation in which the index X_c characterizes the peculiarities of the underlying temperature dynamics from autumn through winter to spring without such a breakdown by intervals. We now combine it with the same kind of unitemized precipitation index, i.e. we take the difference of the east and west precipitations during the whole six months from November to April inclusive (the index E^P_W). As a result we obtain the regression equation:

$$Y = 3.9275 - 5.9873 X_c + 1.4885 B_c + 0.1062 E^P_W; \quad (9)$$

$$R = 0.779$$

The corrected correlation coefficient is definitely higher than the corresponding correlation coefficient of the equation by which the parallelism between the yield and only two indices X_c and B_c was established; for it is 0.752 here against 0.704 there. It is consequently entirely rational to include the criterion-symptom E_{12}^{b2} . The last two equations are the result.

I do not think it is superfluous to note once again the essential difference between them. The equation that includes X_c, E_{12}^{b2} and B_c was set up without detailing the prognostic significance of temperatures and precipitation by the separate time intervals, while the equation with the indices T_c, E_{12}^{b2} and B_c was established by taking into consideration the "active" and "neutral" periods. The corrected correlation coefficients that correspond respectively to the former and the latter equation show a very appreciable difference: 0.752 and 0.830. Even the "coarser" equation, however, makes no slight contribution to the cause of long-term forecasting (though that contribution is still inadequate); for the correlation coefficient is still over 0.70.

14. The index of the difference between Northeast and Southwest precipitation for November, January and April (preceding the harvest) was included in the work as a prognostic criterion. The observations at the Kiev and Odessa stations were used for the Southwest, and those at Kazan' and Ul'yanovsk for the Northeast. The significance of this index, however, is weakened by the fact that owing to the great variety in the manifestations of precipitation, the observations of only two stations are by no means adequate to cover it properly. I therefore worked up the 1883 data for all stations having such data, and available to me, first with the aim of establishing the regional boundaries, even though only

approximately, and secondly to verify and reinforce the conclusions of the work.

The northeast region proved to have approximately the following outlines: its northern boundary runs along the line Molotovsk-Kirov, thence southwest to Gor'kiy-Moscow-Voronezh-Ostrogorsk, thence northeast to Penza, Ul'yanovsk, Polibino, Ufa, Zlatoust, and back to Molotovsk. Within the limits of this region, I have at my disposal observations from the following stations. The Stations shown in italics (underlined in this text) have observations since 1883; the others for shorter periods than that, but in no case less than 30 years.

	r_{y,p_E}		r_{y,p_E}		r_{y,p_E}
<u>Molotovsk</u>	+0.345	<u>Gulynki</u>	+0.210	<u>Ostrogorsk</u>	+0.252
<u>Zlatoust</u>	+0.223	<u>Skopin</u>	+0.220	<u>Ryazan'</u>	+0.118
<u>Gor'kiy</u>	+0.214	<u>Michurinsk</u> (<u>Kozlov</u>)	+0.145	<u>Dankov</u>	+0.166
		<u>Tambov</u>	+0.374	<u>Yelat'ma</u>	+0.173
<u>Kazan'</u>	+0.438	<u>Voronezh</u>	+0.242	<u>Ariashka</u> (<u>Penza</u> <u>Guberniya</u>)	+0.263
<u>Polibino</u>	+0.249	<u>Uryupino</u>	+0.238		
<u>Ul'yanovsk</u>	+0.173	<u>Zemetchino</u>	+0.304	<u>Poret'skoye</u>	+0.251
<u>Kirov (Vyatka)</u>	+0.150	<u>Penza</u>	+0.221	<u>Veshkayma</u> (<u>Ul'yanovskiy</u> <u>Rayon</u>)	+0.335
<u>Moscow</u>	+0.160	<u>Zadonsk</u>	+0.218	<u>Ufa (29 years)</u>	+0.270

Thus all of the 25 stations in the region outlined, for which I had data of observations for 30 years and over, give correlation coefficients with the plus sign.

We compute the mean total precipitation for November, January and April, of each year, at the 16 stations with full observations

over the entire 33 years, and correlate it with the crop yields (the y series). The correlation coefficient attains the value of plus 0.374.

Let us now see what the correlation coefficients between yield and precipitation are for the stations located outside of this region.

Meshecherskoye)	0.074 former	Taganrog	-0.066
Korystovo	-0.010 Tula	Grun' (Poltava guberniya)	0.085
Boloto	0.020 guberniya	Nikolayevskoye (Saratov guberniya)	-0.160
		Voll'sk	0.094
Karachev	-0.142 former	M. Uzen'	0.022
Orel	0.098 Orlov	Ural'sk	0.086
Trosna	0.045 guberniya	Sverdlovsk	0.069
		Arkhangel'sk	-0.070
		Leningrad	-0.021
Akhityrica	0.085 former		
Kupyansk	0.088 Kharkov		
Kharkov	-0.017 guberniya		
Lugansky	-0.075		

If we consider these 19 stations that border the region on the west, south and east, we note in the first place that the signs of the correlation coefficients are not constant. Minus signs alternate with plus signs. Secondly, the coefficients are negligible in size. In 15 cases they are below 0.10, and in two cases (Karachev and Nikolayevskoye, former Saratov guberniya) where they are over 0.10 but under 0.20, they have minus signs instead of plus signs, i.e. they are opposite in sign to the region just considered. But these two stations cannot be included, either, in the other region to the southwest, since they are separated from it by other stations with plus signs.

As we proceed further towards the Southwest, we find ourselves in a region where the November-January-April precipitation has prognostic value, but the reverse of that found for the northeast region, namely: low precipitation during these months is here a favorable

symptom for the coming harvest, while abundant precipitation is unfavorable: all the correlation coefficients have the minus sign:

<u>Kerch</u>	-0.141	<u>Odesa</u>	-0.417
<u>Kherson</u>	-0.087	<u>Kiev</u>	-0.508
<u>Nikolayev</u>		<u>Uman</u>	-0.390
(Kherson guberniya)	-0.128	<u>Aleksayevka</u>	-0.226
Kremenchug	-0.154 former	<u>Korostyanev</u>	-0.528 former
Kobylakha	-0.408 Chernig	<u>Solov'yevka</u>	-0.357 Kiev
Nashin	-0.205 guberniya	<u>Vasilevichi</u>	-0.208 guberniya
Chernigov	-0.375	<u>Pinsk</u>	-0.317
<u>Kirovo</u>			
(Yelisavetgrad)	-0.209		

All the stations with data at my disposal covering observations for 30 years and over were used by me, and stations with the full term of observation (33 years) have been shown in italics (underlined in this text). It will be seen from this data that all 16 stations have negative coefficients of correlation, which are higher than those of the stations in the region to the East (for instance, Grun', Kharkov, Akhtyrka, Kopyansk, Lugansk, Taganrog). Only the "Kherson" station has coefficients below 0.10, but even this station lies at the edge of the region. The approximate eastern boundary of this region is formed by the line of the Dnepr, then the Desna; the boundary then turns west and runs through Vasilevichi towards Pinsk.

We next follow the course of this boundary, using the observations of foreign stations for this purpose. First of all we must observe that the region does not extend all the way to the Atlantic Ocean or the Baltic Sea; here is a list of the stations in Spain, France, Holland, Denmark and North and West Germany, for which I have data for the 33 years:

	Correlation coefficients		Correlation coefficients
Lisbon	0.024	Nantes	-0.040
Madrid	-0.046	Paris	-0.148
Gibraltar	0.023	Frankfurt am Main	-0.052
Palma	-0.062	Utrecht	-0.027
Marseilles	0.067	Kopenhagen	0.071
Lyons	0.030	Konigsberg	-0.036

The correlation coefficients for these stations change their signs and are most negligible in size. Only Paris has a value over -0.10, but even here it is under -0.15.

But the following 13 stations, located in Poland, southeast and south Germany, Austria, Yugoslavia, Italy and Rumania -- on the territory directly adjacent to our own southwest region -- present a different picture both as to signs and values of the correlation coefficients.

Warsaw	-0.323)	Vienna	-0.152)
L'vov	-0.238)	Poland Obir	-0.322)
Czernowits	-0.269)		
		Khvar	-0.295)
Berlin	-0.141)		
Breslau	-0.282)	Germany Milan	-0.101)
Trier	-0.288)	Rome	-0.257)
		Sassari	-0.329)
		Sulina	-0.161)
			Rumania

This table shows that all the correlation coefficients have the same sign and are higher than those of the preceding group of stations. The correlation coefficient between yield and November-January-April precipitation, calculated as the mean value for these 13 stations, is -0.447.

The southwest region thus extends beyond the boundaries of ^{that} the USSR. The line/bounds it runs from Minsk through Warsaw, Berlin,

Trier, Milan, Rome, Sassari, thence turning back to Yugoslavia (Khvar) and Salina, thence to the Crimean peninsula (or the mouths of the Danube), where it meets the southwest region of the USSR. (I can only establish the southern boundary of the region approximately, since I have insufficient data for the South (no observations at all for Bulgaria, Turkey and Greece, and only a single station for Yugoslavia). There is a reason to believe that only a part of Rumania is included within the region; its southern part is not; at least Bucarest (located to the south of Salina) has correlation coefficient of plus 0.009.)

We are now in a position to pass from the characteristics of the difference between northeast and southwest precipitation, prepared from the data of only four stations (Kiev-Odesa and Kasan'-Ul'yanovsk), to the differences between the southwest precipitation, with 22 stations, and the northeast precipitation with 16 stations (38 in all), including 13 stations outside the USSR. As we have seen, the mean precipitation for the 16 stations of the northeast correlates with the crop yield to the extent of plus 0.374, and, for the 22 stations of the southwest, to the extent of -0.517. The difference between northeast and southwest precipitation gives a correlation coefficient of 0.665, which is somewhat higher than that for the four stations included in the work (which is 0.63).

The regression equation for two variables has the following

form:

$$Y = 17.3104 - 3.0345 P_w + 2.8451 P_E; R = 0.665 \quad (10)$$

where P_w is the mean precipitation for 22 stations of the Southwest during November, January and April, while P_E is the mean precipitation during the same months at 16 stations of the Northeast.

CHAPTER V.

15. up to now we have been using multiple regression equations. Their theory, as has been indicated in the Introduction, cannot yet be considered to have been fully worked out. The question thus arises whether all the criterion-symptoms included in this study could not be expressed in a single integral index, thus avoiding the use of multiple regression equations.

9 criteria were selected by us. If aP_w and eP_w are separated into their component elements (O_c and N , P_E and P_W), the number of indices is increased to 11. These indices, with their corresponding correlation coefficients, are as follows (see Table 15 on pages 82-83).

O	-0.344	N	plus 0.372
O-1	-0.379	B	plus 0.263
A	-0.385	B-1	plus 0.267
A-1	-0.286	P_E	- 0.517
J	plus 0.334	P_W	plus 0.387
O_c	-0.404		

As we see, the correlation coefficients do not particularly differ among themselves (except for P_W , which stands out in its size). If it is, moreover, taken into consideration that their mean errors, in view of the small number of observations (33 in all) are fairly large, we may, without serious fault, consider the prognostic value of all these indices to be equivalent. We therefore define the general integral index as the mean value of the 11 cri-

teria enumerated. In deriving this mean, however, the following points must be kept in mind: first, the values O , O_{-1} , A , A_{-1} , etc. down to P_E and P_W inclusive, are qualitatively different; some of them fix the temperature, others the time of unfolding the leaves, still others the precipitation, consequently their simple summation is impossible. However, if each series is expressed in the form of standard deviations (i.e. in "signas") (these "signas" are given in Table 13 for each index), then this qualitative difference is eliminated and their summation becomes entirely legitimate.

The series O , O_{-1} , A , A_{-1} , etc. down to and including P_E and P_W , must first be transformed into the series: $\frac{O}{O}$, $\frac{O_{-1}}{O_{-1}}$, etc. down to $\frac{P_E}{P_E}$, $\frac{P_W}{P_W}$. Secondly, the arithmetic means of the 11 indices must be based on their algebraic sum, since the correlation coefficients corresponding to them have different signs. As a result the integral index, which we shall designate by the symbol Θ_{11} , is defined by the following formula: $\Theta_{11} =$

$$\frac{1}{11} \left(\frac{O}{O} + \frac{N}{3N} + \frac{E}{E} + \frac{B_{-1}}{B_{-1}} + \frac{O}{O} + \frac{O}{O} + \frac{O_{-1}}{O_{-1}} + \frac{A}{A} + \frac{A_{-1}}{A_{-1}} + \frac{E_{-1}}{E_{-1}} + \frac{P_E}{P_E} \right)$$

The values of the index Θ_{11} were calculated by this formula for each year, and are given in Table 18. The correlation coefficient between the yield (y) and the integral index Θ_{11} is 0.867, while the equation of regression between y and Θ_{11} has the following form:

$$Q = 31.1 + 2.5847 Q_1, R = 0.867 \quad (10)$$

(11)
(See footnote below)

[See Table 18 on following page]

(It is very easy to prove mathematically that both the correlation coefficient (0.867) and the value of $1Y$ are entirely identical with the results that would be yielded by the calculation if we had used the values of the deviations, expressed in sigmas, of each of the 11 indices from its mean, to derive the integral index; i.e. if we have based it on the values

$$\frac{Q - Q_1}{30}, \frac{Q_2 - Q_1}{30}, \text{ etc.}$$

instead of on the values

$$\frac{Q}{30}, \frac{Q_1}{30}, \text{ etc.} \quad \text{or} \quad \frac{P}{30}$$

In this case, both integral index Q_{11} and the regression equation would have assumed a different form, but the same values for R and $1Y$ would have been obtained.)

Column 3 shows the values for $1Y$ calculated by this equation while column 4 shows the divergence between these values and the actual crop yields. It will be seen that the maximum divergence occurred in 1915 and was somewhat less than 15 percent (14.8 percent). Of the 33 cases, only 6 exceed 10 percent (1885, 1893, 1894, 1897, 1914, 1915); in 9 cases the "forecast" was for a yield more than 7 percent below the average, and the actual yield for all these years

TABLE 18

Year	y	ϕ_{11}	y^*	$y - y^*$	Year	y	ϕ_{11}	y^*	$y - y^*$
1883	+0.5	-0.44	-0.2	+0.7	1900	+2.5	-0.31	+3.5	-1.0
1884	+7.0	-0.13	+8.6	-1.6	1901	-13.9	-0.73	-8.5	-5.4
1885	-10.8	-0.37	+1.8	-12.6	1902	+15.3	+0.02	+12.9	+2.4
1886	0.0	-0.35	+2.3	-2.3	1903	+4.0	+0.03	+13.2	-9.2
1887	+14.6	+0.16	+16.9	-2.3	1904	+18.9	+0.01	+12.6	+6.3
1888	+9.1	-0.23	+5.7	+3.4	1905	-7.4	-0.37	+1.8	-9.2
1889	-18.2	-1.01	-16.5	-1.7	1906	-23.9	-1.24	-23.0	-0.9
1890	-7.3	-0.01	-13.6	+6.3	1907	-10.0	-1.05	-17.6	+7.6
1891	-29.0	-1.60	-33.3	+4.3	1908	-8.5	-0.83	-11.4	+2.9
1892	-13.2	-0.69	-7.4	-5.8	1909	+15.7	+0.09	+14.9	+0.8
1893	+21.8	-0.09	+9.8	+12.0	1910	+6.4	-0.32	+3.2	+3.2
1894	+22.0	-0.04	+11.2	+10.8	1911	-18.2	-0.85	-11.9	-6.3
1895	+9.0	-0.18	+7.2	+1.8	1912	+7.9	-0.20	+6.6	+1.3
1896	+4.8	-0.29	+4.0	+0.8	1913	+17.1	-0.12	+8.9	+8.2
1897	-14.3	-0.53	-2.9	-11.5	1914	-11.5	-0.41	+0.6	-12.1
1898	-1.2	-0.37	+1.8	-3.0	1915	+10.9	-0.57	-3.9	+14.8
1899	+10.0	+0.05	+13.7	-3.7					

proved below the average (1891, 1906, 1889, 1911, 1901, 1907, 1892, 1890, 1908); in 16 cases the "forecast" predicted a yield more than 3 percent over average, and all these 16 years were marked by a yield above the average.

[See diagram 6 on following page]

Diagram 6 graphically demonstrates the strong parallelism between the values of y calculated by the regression equation ($y = 12.3116 + 28.5087 \phi_{11}$), and the actual y vals. As on Diagram 5, the calculated values of y are plotted along the axes of abscissae, while the actual indice of the height of the yields are plotted along the axes of ordinates.

DIAGRAM 6

COMPARISON OF THE OBSERVED YIELDS WITH THOSE CALCULATED
BY THE REGRESSION EQUATION:

$$y = 12.3116 + 28.5067 x_{11}$$

+23

+21

+19

+17

+15

+13

+11

+9

+7

+5

+3

+1

-1

-3

-5

-7

-9

-11

-13

-15

-17

-19

-21

-23

-25

-27

Yields calculated by the regression equation: $y = 12.3116 + 28.5067$

Caption of

Base Map 2 (Page 86)

COEFFICIENT OF CORRELATION OF CROP YIELD WITH THE
TEMPERATURE OF THE FIRST OCTOBER DECADE

(r_{y, o_1})

Caption of

Base Map 3 (Page 87)

COEFFICIENT OF CORRELATION OF CROP YIELD WITH THE
TEMPERATURE OF THE FIRST OCTOBER DECADE

(r_{y, A_3})

Caption of

Base Map 4 (Page 88)

COEFFICIENT OF CORRELATION OF CROP YIELD WITH THE
TEMPERATURE OF THE FIRST OCTOBER DECADE
OF THE PRIOR AGRICULTURAL YEAR

$(r_{y, o-1})$

Caption of

Base Map 5

CORRELATION COEFFICIENTS OF CROP YIELD WITH THE
TEMPERATURE OF THE THIRD APRIL DECADE
OF THE PRIOR AGRICULTURAL YEAR

$(r_{y, A-3})$

CHAPTER VI

16. It has been indicated in the Introduction that the empirical regularities established were partially verified. In what did this verification consist, and what were its results.

The 33-year period was first divided into two intervals: the first from 1883 to 1899 (17 years), and the second from 1900 to 1915 (16 years). Each of these intervals covers a sufficiently varied group of harvests. Each of them includes good harvests, near-average harvests, bad ones. The algebraic sum of the deviation over and under is equal in the first case to plus 5.5 and in the second case to plus 5.3, with the total over-average yields almost coinciding with the total under-average yield (1883-1899: 99.6 and -94.0; 1900-1915: 98.7 and 93.4). All this shows the two periods to be suitable for the verification.

For each of these two periods the integral index was determined on the basis of the 11 series (O , O_{-1} , A , A_{-1} , J , O_c , H , B , B_{-1} , I , P_a , P_e), and of the sigmas corresponding to each period, in precisely the same way as this was done for the whole of the 3-year period. For the period 1883-1899 this index, which we shall designate by the symbol $17O_{11}$, has the following values.

[Table 19 shown on following page]

The correlation coefficient between the crop yield (y) and the index O_{11} has the value 0.866, while the equation of regression between y and $17O_{11}$ has the following form: [$*_{17O_{11}}$ intended]
 $y_{17} = 12,02571 + 26.09307 17O_{11}$ $R = 0.866$.

TABLE 19

Year	17 11	1 ^Y 17	Y	Year	17 11	1 ^Y 17	Y
1883	-0.42	+1.1	+0.5	1892	-0.73	-7.0	-13.2
1884	-0.12	+8.9	+7.0	1893	-0.11	+9.2	+21.8
1885	-0.34	+3.1	-10.8	1894	-0.05	+10.7	+22.0
1886	-0.36	+2.6	0.0	1895	-0.20	+6.8	+9.8
1887	+0.13	+15.4	+14.6	1896	-0.28	+4.7	+4.8
1888	-0.29	+4.5	+9.1	1897	-0.55	-2.3	-14.3
1889	-1.10	-16.7	-16.2	1898	-0.42	+1.1	-1.2
1890	-1.07	-15.9	-7.3	1899	+0.01	+12.3	+10.0
1891	-1.72	-32.9	-29.0				

Column 2 gives the numerical values of 17_{17} as determined by this equation.

We next apply this equation to the second period, covering the years 1900-1915 inclusive. Attention is called to the fact that these observations were not used to derive this regression equation (12). Thus we extrapolate this regression equation to the years 1900-1915. For the extrapolation we must first find the integral index for each year from 1900 to 1915. The observation for these years for the series O, O_{-1} etc. were given in table 15. We then put them in the form $\frac{O}{O_0}, \frac{O_{-1}}{O_{0-1}}$, etc. using the signs of the preceding period for this purpose, and, finally, take the arithmetic mean of these series, which we denote by the symbol 17_{011} . The calculated values of 17_{011} for 1900-1915 are given in the following table:

[See Table 20 on following page]

The extrapolation is concluded by the substitution in (12), which is based on data from the other period (1883-1899), of the values of 17_{011} for each year of the period 1900-1915, in order to determine the crop yields to be expected on the bases of the criteria O, O_{-1} , etc. through P_n inclusive. The indices of expected yield so calculated are presented in column 2 of the table under the heading $y_{extraneous}$ (which is intended to mean that these indices have been calculated according to an "extraneous" equation, based on data of another period). It is clear from all this that this process coincides with that used for forecasts: the regularities established from the data of one interval are applied to years not

TABLE 20

Year	17	11	$y(y)$	y	$-y(y)$	Year	17	11	$y(y)$	y	$y-y(y)$
1900	-0.31		+3.9	+2.5	-1.4	1908	-0.97		-13.3	-8.5	-4.8
1901	-0.87		-10.7	-13.9	-3.2	1909	+0.02		+12.5	+15.7	+3.2
1902	-0.08		+9.9	+15.3	+5.4	1910	-0.44		+0.5	+6.4	+5.9
1903	-0.07		+10.2	+4.0	-6.2	1911	-1.05		-15.4	-16.2	-2.8
1904	-0.19		+7.1	+18.9	+11.8	1912	-0.24		+5.8	+7.9	+2.1
1905	-0.54		-2.1	-7.4	-5.3	1913	-0.15		-8.1	+17.1	+9.0
1906	-1.52		-27.6	-23.9	+3.7	1914	-0.54		-2.1	-11.5	-9.4
1907	-1.32		-22.4	-10.0	+12.4	1915	-0.63		-4.4	-10.9	-15.3

included in that period. Let us see how far the regularities of the 1883-1899 period, expressed in regression equation (12), are justified during the course of the following 16-year period (1900-1915). For this purpose we use that formula for the correlation efficient which expresses the degree of coincidence of the values of y calculated by the regression equation with the actually observed values y . As is well known, this formula has the following form:

$$R = \frac{\sum (y - y_0)^2 - \sum (y - y^1)^2}{\sum (y - y_0)^2}$$

where y_0 is the arithmetic mean of the series y , and y^1 that of the value determined by the "proper" [own period] regression equation. When we extrapolate the regression equation to another time interval, we take the values of $y_{\text{extraneous}}$ instead of y^1 ; of course y and y_0 likewise relate to the interval to which we are extrapolating. As a result of the preceding formula is transformed to the following:

$$P = \frac{\sum (y - y_0)^2 - \sum (y - y_{\text{ext}})^2}{\sum (y - y_0)^2}$$

If $P = 0$, we must recognize the system of forecasting to be unconditionally ineffective.

Rho P may in fact vanish only on condition that

$\sum (y - y_{\text{ext}})^2 = \sum (y - y_0)^2$. This equation would show that the accuracy of the forecasting system being tested was not at all

126-101

better than that of forecasts made by using the arithmetic mean, for the sum of the squares of the deviations of the actual values y from those expected according to forecasts is equal to the sum of the squares of the deviations of those same values y from their mean. If ρ is greater than 0, (which is possible only on condition that the actually observed values of y are closer to the values $y_{\text{extraneous}}$ than to the arithmetic mean y_0), namely, if $\sum (y - y_{\text{ext}})^2 < \sum (y - y_0)^2$, then the forecasting system has a real foundation, and the more so, the higher the value of ρ .

Finally, if $\sum (y - y_{\text{ext}})^2 = 0$, which makes ρ equal to unity, we have complete coincidence between the predicted values $y_{\text{extraneous}}$ and the actually observed values y (the case of ideal full justification of the forecasting system) (It can easily be seen that the third possibility $\sum (y - y_{\text{ext}})^2 > \sum (y - y_0)^2$ would testify to an exceptional degree of unsuccessfulness in the forecasting system. The application of this method of determining the degree of justification of forecasts was given by me in my published paper: V. M. Obukhov "The implications of the phenological observations of I. A. Pul'man at the Bogoroditskoye-Fenino Meteorological station, in the Central Chernozem Region", 1933, Zhurnal Geofiziki (Journal of Geophysics), vol. III, fascicle 2.) We now apply this formula to the proposed forecasting system (namely, to the extrapolation of the equation set up on the observations of 1883-1899, to the period 1900-1915). As a result we have $\rho = 0.826$. The size of the coefficient ρ is most definitely in favor of the proposed system of forecasting by the integral index 17011 composed of 11 criteria-symptoms. It must also be emphasized that the coefficient ρ , equal to 0.826, is very close to the correlation co-

74 / 09

efficient between 17011 and the 1883-1899 yields, which is 0.866.

We proceed in exactly the same way with the regression equation set up on the basis of the observation for 1900-1915, and extrapolate it to the period 1889-1899. The integral index, designated by the symbol 16^{011} , correlation with the 1900-1915 yields to the extent of 0.840, and the regression equation for this period is as follows:

$$1716 = 17.67865 + 27.67209 \ 16^{011}, \ R = 0.840 \quad (13)$$

We apply this equation to the interval 1883-1899. For this we first set up the series 16^{011} for the interval 1883-1899. This series is based on the observations of the values of 0, 0-1, A, A-1, etc. through P_w inclusive of this interval, and on the signs of the 1900-1915 period. The values of rho are then substituted in the equation just derived, and as a result y ext is found for each year of the 1883-1899 period. These indices are given in the following table:

[See table 21 on following page]

We determine the value of the coefficient rho for 1883-1899, in exactly the same way as was done for the 1900-1915 period, and find it to be 0.885. This is very definitely evidence in favor of the proposed system of forecasting by the use of the integral index based on the 11 criteria-symptoms.

Thus the verification of this system, made by dividing the main period 1883-1915 into two intervals and using the regression

TABLE 21

Year	16	11	$y(y)$	y	$y-y(y)$	Year	16	11	$y(y)$	y	$y-y(y)$
1883	-0.74	-2.8	+0.5	+3.3	1892	-1.01	-10.3	-13.2	-2.9		
1884	-0.35	-8.0	+7.0	-1.0	1893	-0.33	+6.5	+21.8	+13.3		
1885	-0.57	+1.9	-10.8	-12.7	1894	-0.22	+11.6	+22.0	+10.4		
1886	-0.57	+1.9	0.0	-1.9	1895	-0.37	+7.4	+9.8	+2.4		
1887	+0.06	+16.0	+14.6	-1.4	1896	-0.57	+1.9	4.8	+2.9		
1888	-0.44	+5.5	+9.1	+3.6	1897	-0.80	-4.5	-14.3	-9.8		
1889	-1.31	-18.6	-18.2	+0.4	1898	-0.56	+2.2	-1.2	-3.4		
1890	-1.16	-14.4	-7.3	-7.1	1899	-0.16	+13.2	+10.0	-3.2		
1891	-1.89	-34.6	-29.0	+5.6							

equation of the first interval as a forecasting equation for the second interval, and vice versa, using the regression equation of the second interval as forecast of the first one, gave results which were eminently satisfactory: rho not only has a very high value, but, in addition, holds to approximately the same level for both "proper" and "extraneous" equations. The indices are as follows:

PERIODS	"Proper" equation (coefficient R)	"Extraneous" equation (coefficient p (rho))
1883-1899	0.866	0.885
1900-1915	0.840	0.826

The closeness of these coefficients to each other compels us to assume that as strong parallelism should be observed between the numerical values of the indices of the crop yield to be predicted according to its own "proper" equation and to the "extraneous" equation.

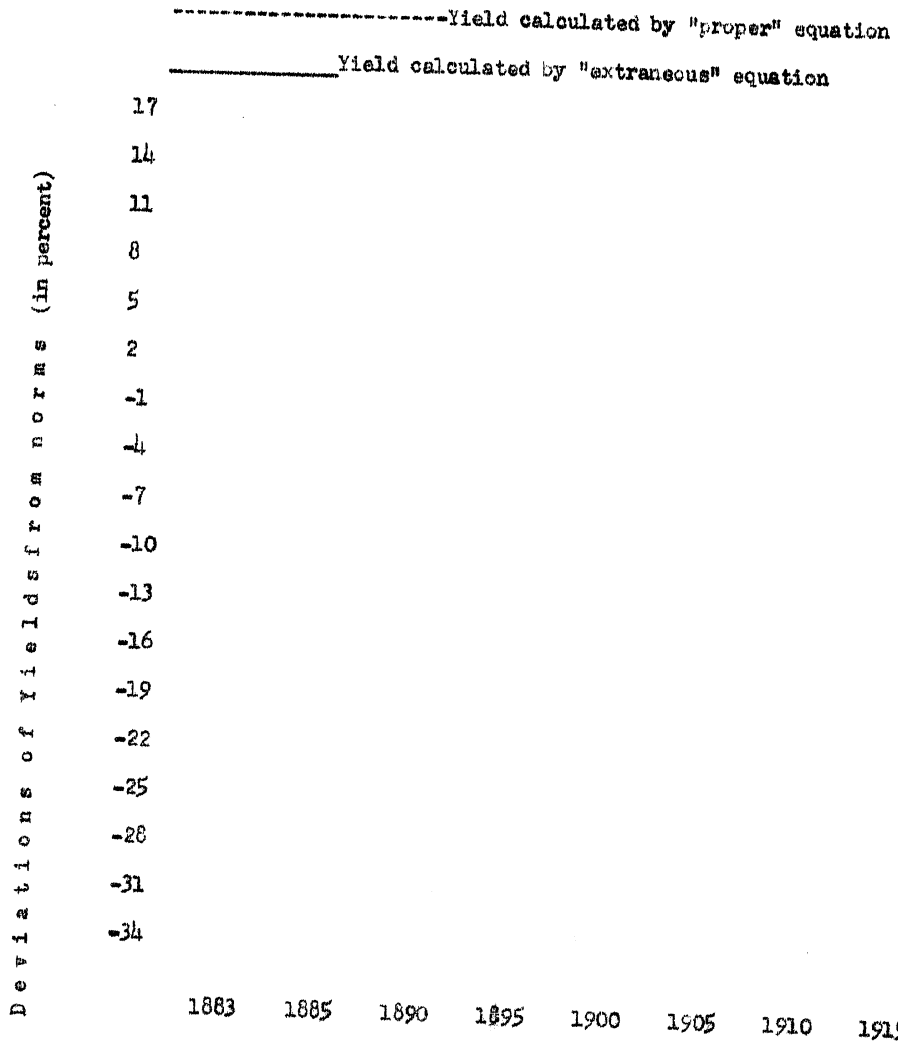
[See Diagram 7 on following page]

Table 22 gives the results of these calculations, one beside the other. The correlation coefficient between the values " y_{pr} " determined by the "proper" equation and the values " y_{ext} " attains the value of 0.986 [sic], which evidences an extremely close parallelism which is only slightly less than complete coincidence of these series.

27/12

DIAGRAM 7

YIELD INDICES CALCULATED BY "PROPER" AND EXTRANEEOUS" EQUATIONS



78-3

Table 22

Year	y_{ext}	y_{pr}	Year	y_{ext}	y_{pr}
1883	-2.8	+1.1	1900	+3.9	+2.7
1884	+8.0	+8.9	1901	-10.7	-7.8
1885	+1.9	+3.1	1902	+9.9	+13.0
1886	+1.9	+2.6	1903	+10.2	+13.3
1887	+16.0	+15.4	1904	+7.1	+13.8
1888	+5.5	+4.5	1905	-2.1	+2.2
1889	-18.6	-16.7	1906	-27.6	-22.2
1890	-14.4	-15.9	1907	-22.4	-17.2
1891	-34.6	-32.9	1908	-13.3	-11.1
1892	-10.3	-7.0	1909	+12.5	+14.4
1893	+8.5	+9.2	1910	+0.5	+3.6
1894	+11.6	+10.7	1911	-15.4	-11.9
1895	+7.4	+6.8	1912	+5.8	+6.1
1896	+1.9	+4.7	1913	+8.1	+8.3
1897	-4.5	-2.3	1914	-2.1	+1.9
1898	+2.2	+1.1	1915	-4.4	-3.6
1899	+13.3	+12.3			

r

0.986 (see Diagram 7)

79

CHAPTER VII

17. We shall now verify the proposed forecasting system by another method.

We have established a very considerable parallelism between the composite index of 11 criteria-symptoms and the yield. This parallelism is measured by the correlation coefficient 0.867. These component indices, which characterize the weather of late autumn, winter and early spring, can of course exert no direct influence upon the crop. The parallelism noted can only occur because these criteria are the precursors of one course or another for the spring and summer weather which is reflected in the upward or downward fluctuations of the yield. If this is really the case, then a correlation should be observed between the index. O_{11} and the basic meteorological elements that characterize the humidity of the weather. We carried out such a verification with respect to the precipitation and the temperature at 1300 hours. The regions for which these weather indices were tabulated embrace those guberniyas of pre-revolutionary Russia for which the proposed forecasting system proved effective (see Base Map 1). The period from 1886 to 1915 inclusive was taken for the verification, since during the first three years, 1883, 1884 and 1885, the number of stations was insufficient to characterize the precipitation. The tabulation used data of stations with uninterrupted service and of those where breaks in service could be filled in by data from other near-by stations. The interpolation for such breaks was done according to the differences between the mean precipitation, based on long-term records, at the base and auxiliary stations.

This is the reason why the precipitation data for each year are given in Table 23 in terms of departures from the long-term means. In this table the symbols P_V , P_{VI} and P_{VII} denote the precipitation during May, June and July respectively, while T_V , T_{VI} and T_{VII} denote the mean temperature at 1300 hours for those months respectively. The coefficients of correlation between the integral index and the series P_V , P_{VI} and P_{VII} and T_V , T_{VI} and T_{VII} , have the following values:

for the May precipitation	0.509	for the May temperature	-0.391
for the June precipitation	0.450	for the June temperature	-0.433
for the July precipitation	0.230	for the July temperature	-0.596

Thus the signs we have noted really turn out to be heralds of the character of the spring and summer weather: the higher (algebraically) the indices of Q_{11} , the higher the rainfall and the lower the temperature to be expected, while on the contrary arid years correspond to (algebraically) lowered Q_{11} indices: that is, years with scanty rainfall and high temperatures. The "forecasting" significance of Q_{11} is augmented for temperature increases with them and reaches its peak effect in July, while for precipitation it falls from May to June, and falls especially from June to July. It should be noted that the significance of the July rainfall for the yield (as shown by a study of the dependence of yield on meteorological factors) is negligible, while the influence of temperature still continues to be marked. For this reason the July rainfall may be disregarded in deriving the general index of the character of the spring and summer weather (insofar as this interests us in its influence on the yield). This general index is shown in the table under the symbol V_5 . It is based on May and June rainfall and

8176

May, June and July temperature.

To derive it, the series P_V , P_{VI} , T_V , T_{VI} , T_{VII} were expressed as the "sigmas" of each series and are entered in the table under the symbols

$$\frac{P_V}{\Omega}, \frac{P_{VI}}{\Omega}, \frac{T_V}{\Omega}, \frac{T_{VI}}{\Omega}, \frac{T_{VII}}{\Omega}$$

V_5 is computed according to the formula: $V_5 = \frac{1}{5} \left(\frac{P_V}{\Omega} + \frac{P_{VI}}{\Omega} + \frac{T_V}{\Omega} + \frac{T_{VI}}{\Omega} + \frac{T_{VII}}{\Omega} \right)$

Since the correlation between precipitation and temperature is negative (the higher the temperature, the lower the precipitation, and conversely the higher the precipitation, the lower the temperature), and since drought is characterized by increased temperatures and scanty rainfall, the indices P_V and P_{VI} are thus taken with plus signs and indices T_V , T_{VI} and T_{VII} with minus signs.

Let us see now what correlation exists between O_{11} and the general index of the character of the spring and summer weather, V_5 . The correlation coefficient between these series is equal to:

$$r_{O_{11}, V_5} = 0.712$$

This coefficient is high enough to allow us to state very definitely that the peculiarities we have pointed out in the march of the weather from the end of autumn through the winter to the spring serve as signals for those processes taking place in the atmosphere which are decisive in spring and summer either for aridity or humidity. It is, of course, impossible to consider that these

22, 1/7

criterion-symptoms exhaust the problem of forecasting the character of the spring and summer weather, for the correlation coefficient has not attained unity, but we can recognize it as half solved, and this is no small contribution to the problem of long-range forecasting, which is so difficult and so important. And the size of the coefficient of correlation between Θ_{11} and V_5 does give us the right to assure that the problem has been half solved.

Indeed, if any other investigation, employing other criterion-symptoms, not correlatively linked with ours, succeeded in obtaining such a result, measured by a correlation coefficient of such a magnitude (approximately 0.70 to 0.71), then both of these investigations - the one herein offered and the hypothetical one -- would have completely solved the problem of forecasting the spring and summer weather. The joint employment of both would yield ideal forecasting accuracy, that is, the complete coincidence of the values for B_5 obtained by observation with the forecast made in accordance with the regression equation:

$V_5 = a + b\Theta_{11} + c\Theta_{hyp}$, in which Θ_{11} is our own integral index and Θ_{hyp} is the integral index of the hypothetical investigation.

18. Finally, I set myself the task of verifying, by still a third method, the extent to which the hypotheses herein established correspond with reality. It is possible to attain conviction on this score by means of extending the results obtained to new periods, not covered by the investigation proper. With the materials I succeeded in collecting, I was enabled to do this with respect to the periods 1872-1882 and 1920-1931 (the original investigation having covered the period 1883-1915)

84-115

Table 23

YEAR	P _V	P _{VI}	P _{VII}	T _V	T _{VI}	T _{VII}	O ₁₁	$\frac{P_V}{O_V}$	$\frac{P_{VI}}{O_{VI}}$	$\frac{T_V}{O_V}$	$\frac{T_{VI}}{O_{VI}}$	$\frac{T_{VII}}{O_{VII}}$	$\frac{V_S}{(1h)}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1886	+11.5	+19.0	+28.3	17.8	22.0	23.4	-0.35	+1.00	+1.29	7.95	13.13	14.07	-6.57
1887	-3.0	+4.0	-4.3	20.7	20.6	23.3	+0.16	-0.26	+0.27	9.24	12.29	14.01	-7.11
1888	+3.5	-8.3	+27.6	18.5	20.3	24.3	-0.23	+0.31	-0.56	8.26	12.11	14.61	-7.05
1889	-15.8	-0.5	-20.8	20.4	21.6	26.1	-1.01	-1.38	-0.03	9.11	12.89	15.69	-7.82
1890	-12.8	+20.3	-16.2	18.4	23.5	28.1	-0.91	-1.12	+1.38	8.21	14.02	16.90	-7.77
1891	-22.5	-13.7	-26.8	20.1	23.8	28.8	-1.60	-1.96	-0.93	8.97	14.20	17.32	-8.68
1892	-8.5	-15.6	-8.4	19.7	24.2	26.1	-0.69	-0.74	-1.06	8.79	14.44	15.69	-8.14
1893	-13.4	+13.0	+5.1	16.1	22.1	25.7	-0.09	-1.17	+0.88	7.19	13.19	14.25	-6.98
1894	+3.3	+28.9	-5.9	19.2	19.6	22.9	-0.04	+0.29	+1.96	8.57	11.69	13.77	-6.36
1895	-4.0	-9.0	-14.1	15.9	21.9	25.3	-0.18	-0.35	-0.61	7.10	13.07	15.21	-7.27
1896	+11.8	+4.0	+28.6	17.1	21.4	22.6	-0.29	+1.03	+0.27	7.63	12.77	13.59	-6.54
1897	+4.2	-18.4	-10.8	22.0	23.3	26.1	-0.53	+0.37	-1.25	9.82	13.90	15.69	-8.06
1898	-8.4	-8.2	+6.5	21.2	21.9	26.7	-0.37	-0.73	-0.56	9.46	13.07	16.06	-7.98
1899	-10.2	+6.8	-13.0	17.5	21.0	24.6	+0.05	-0.89	+0.46	7.81	12.53	14.79	-7.11
1900	+2.3	+16.2	-0.6	17.6	20.0	23.7	-0.31	+0.20	+1.10	7.86	11.93	14.25	-6.55
1901	+7.3	-36.7	-19.5	17.7	26.5	24.9	-0.73	+0.64	-2.49	7.90	15.81	14.97	-8.11

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	1902	+13.2	+ 5.2	+ 4.5	17.0	23.4	24.6	+0.02	+1.15	+0.35	7.59	13.96	14.79	-6.97
	1903	+24.8	+ 2.0	-11.2	16.7	24.0	26.5	+0.03	+2.16	+0.14	7.46	14.32	15.94	-7.08
	1904	+ 1.2	+ 6.4	- 6.7	16.1	19.0	22.8	+0.01	+0.10	+0.43	7.19	11.34	13.71	-6.34
	1905	-14.9	- 7.1	- 5.2	20.6	23.6	23.5	-0.37	-1.30	-0.48	9.20	14.08	14.13	-7.84
	1906	- 7.1	- 8.7	+18.2	23.9	24.2	26.9	-1.24	-0.62	-0.59	10.67	14.44	16.18	-8.50
	1907	- 5.9	-8.8	+8.3	16.6	22.2	25.9	-1.05	-0.51	-0.60	7.41	13.25	15.57	-7.47
	1908	+ 5.5	-0.5	+12.7	14.6	22.8	23.4	-0.83	+0.48	-0.03	6.52	13.60	14.07	-6.75
	1909	+ 8.4	+24.9	-2.4	16.2	22.3	24.1	-0.09	+0.73	+1.69	7.23	13.31	14.49	-6.52
	1910	+ 5.5	+10.5	+ 8.9	18.7	22.0	25.8	-0.32	+0.48	-0.71	8.35	13.13	15.51	-7.44
	1911	-15.3	-11.4	+ 4.5	19.0	22.1	24.8	-0.85	-1.34	-0.77	8.48	13.19	14.91	-7.74
	1912	+25.6	+ 4.9	- 5.4	15.0	24.9	21.4	-0.20	+2.18	+0.33	6.70	14.86	12.87	-6.38
	1913	+ 5.6	-16.1	+48.2	14.1	19.6	23.2	-0.12	+0.49	+1.09	6.29	11.69	13.95	-6.07
	1914	-0.3	-17.3	-17.1	19.3	21.7	25.5	-0.41	-0.03	-1.17	8.62	12.95	15.33	-7.62
	1915	+ 8.7	-19.5	+12.5	16.8	21.6	24.9	-0.57	+0.76	-1.32	7.50	12.89	14.97	-7.18

97
20

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1902	+13.2	+ 5.2	+ 4.5	17.0	23.4	24.6	+0.02	+1.15	+0.35	7.59	13.96	14.79	-6.97
1903	+24.8	+ 2.0	-11.2	16.7	24.0	26.5	+0.03	+2.16	+0.14	7.46	14.32	15.94	-7.08
1904	+ 1.2	+ 6.4	- 6.7	16.1	19.0	22.8	+0.01	+0.10	+0.43	7.19	11.34	13.71	-6.34
1905	-14.9	- 7.1	- 5.2	20.6	23.6	23.5	-0.37	-1.30	-0.48	9.20	14.08	14.13	-7.84
1906	- 7.1	- 8.7	+18.2	23.9	24.2	26.9	-1.24	-0.62	-0.59	10.67	14.44	16.18	-8.50
1907	- 5.9	-8.8	+8.3	16.6	22.2	25.9	-1.05	-0.51	-0.60	7.41	13.25	15.57	-7.47
1908	+ 5.5	-0.5	+12.7	14.6	22.8	23.4	-0.83	+0.48	-0.03	6.52	13.60	14.07	-6.75
1909	+ 8.4	+24.9	-2.4	16.2	22.3	24.1	-0.09	+0.73	+1.69	7.23	13.31	14.49	-6.52
1910	+ 5.5	+10.5	+ 8.9	18.7	22.0	25.8	-0.32	+0.46	-0.71	8.35	13.13	15.51	-7.44
1911	-15.3	-11.4	+ 4.5	19.0	22.1	24.8	-0.85	-1.34	-0.77	8.48	13.19	14.91	-7.74
1912	+25.6	+ 4.9	- 5.4	15.0	24.9	21.4	-0.20	+2.18	+0.33	6.70	14.86	12.87	-6.38
1913	+ 5.6	-16.1	+18.2	14.1	19.6	23.2	-0.12	+0.49	+1.09	6.29	11.69	13.95	-6.07
1914	-0.3	-17.3	-17.1	19.3	21.7	25.5	-0.41	-0.03	-1.17	8.62	12.95	15.33	-7.62
1915	+ 8.7	-19.5	+12.5	16.8	21.6	24.9	-0.57	+0.76	-1.32	7.50	12.89	14.97	-7.18

(I took the data on the 1872-1882 crop yields from the work of the statistician V. G. Mikhaylovskiy "Crop yields in Russia, 1801-1914" (Bulletin No 50, TsSU). This material is repeated in the work of M. I. Semenov "On the question of the regularity in crop yield fluctuations", Vestnik Statistiki (Statistical Messenger), vol. XI. The series of Mikhaylovskiy was fitted by Semenov in a manner analogous to my own procedure for the 1883-1915 period (see Page 49). I utilized this alignment to calculate the percentage ratios of the observed yields to the corresponding normal yield for each year. These indices are given in column 1 of table 25. (As in text).

The yield data for 1920-1931 were taken from the pamphlet "Crop results for 1933". I took this series as my base, with one minor correction: I excluded from the 1924-1925, amounting to 1.7 puds per desyatin, and in this manner restored the original 1924 crop estimate according to the data of the unmodified budget studies. Instead of 41.2 pud per desyatin I thus took 39.5 puds per desyatin. to disclose the part played by the secular trend in the dynamics of crop yields since 1920, I fitted a curve to the data. I founded the work on the following thesis. In consequence of the imperialistic war and the civil war, of the crop delivery quotas and the catastrophic crop failure of 1921, a decline in productivity took place, lasting until the end of 1922. Productivity commenced to increase only in 1923. It is customarily but erroneously assumed that this rise in productivity started in 1922. This is incorrect, for the disruption caused by 1921 (the clashing of livestock supplies, and their exhaustion, the enfeeblement of human working force, the extremely poor quality of seed used for the 1922 harvest)

The following are the series and norms established by me on the basis of that trend fitting:

YEAR	Yield in tsentner per hectare	Norm for yield	Yield in percent of norm
1920	5.69	7.10	80.1
1921	5.04	7.02	71.8
1922	8.06	6.94	116.1
1923	7.12	7.02	101.4
1924	5.92	7.10	83.4
(with my correction)			
1925	8.31	7.18	115.7
1926	8.16	7.27	112.2
1927	7.62	7.38	103.5
1928	7.89	7.44	106.0
1929	7.42	7.52	98.7
1930	8.48	7.60	111.6
1931	6.58	7.68	85.7

(a) First I verified the index T_6 (on this index see Chapter II, pp. 74-80, and the note to Table 25, page 116).

In the investigation proper (covering the 1883-1915 period) this index was computed for Ul'yanovsk station, but it was also mentioned there that we had obtained almost identical results at "Kazan'" station, and in fact the correlation coefficient between y (crop yield) and the index T_6 for the period 1883-1915 is -0.754 at Ul'yanovsk station and -0.723 at Kazan' station. As we see, these correlation coefficients differ very little. I had the subsequent opportunity to work up the 1872-1882 and 1920-1931 data for Kazan'

station, and to find the value of the index T_6 for each year. We thus have a total period of 56 years. Let us divide this period into five short 11-year intervals (12 years for the last interval, covering the post-revolutionary years), and calculate the coefficient of correlation between the yield and the index T_6 .

As a result we obtain:

1872-1882	11 years	r_y, T_6	(As in text,) -0.167 (this period was not included in the investigation proper).
1883-1893	11 years	" "	-0.772
1894-1904	11 years	" "	-0.799 periods of investigation proper
1905-1915	11 years	" "	-0.533
1920-1931	12 years	" "	-0.633 (this period not included in investigation proper)
For all 56 years	" "	" "	-0.639

Thus the coefficients of correlation for each short period all have the minus sign, and consequently the direction of the parallelism between the yield and the index T_6 remains the same, namely; the higher T_6 , the more unfavorable the prospects for the crop yield in the coming year: this minus sign is still maintained through two periods not included in the investigation proper, and in this respect confirms the results of that investigation. Quantitatively, however, the correlation coefficients for these short periods are markedly dissimilar, which indicates that the processing of data for short periods does not yield reliable indices for forecasting purposes.

(Specifically, the low correlation coefficient for 1872-1882 is explained by the low representation of the extreme variants of crop yields throughout this period. In fact the mean first-power deviation [the mean deviation] of the yields about the "norms" for this period = 7.5, and for the remaining periods amounts respectively to 12.0; 10.5; 12.3 and 12.2.)

The combination of these short periods into longer ones gives us considerably more stable correlation coefficients.

Let us divide the whole 44-year pre-revolutionary period into two 22-year periods, and, after setting up the post-revolutionary period separately, compute the correlation coefficients:

		r_y, T_6
1872-1893	22 years	-0.632
1894-1915	22 years	-0.627
1920-1931	12 years	-0.633
A 1	56 years	-0.639

The verification of the forecasting value for the post-revolutionary periods of T_6 , established for the pre-revolutionary period 1883-1915, is of particular interest for us. We therefore show on the Diagram 8 the crop yields from 1920 to 1931, and the indices T_6 . This diagram shows with the utmost clarity: all of the 5 years with crop yields that were reduced, under average, poor or very poor have the highest T_6 indices. On the other hand, the lowest value of T_6 correspond to the 7 years with over-average or good crop yields. Thus the verification of the forecasting value of T_6 , established by a study of the pre-revolutionary epoch, con-

firms the hypothesis for the period 1920-1931 as well (see page 112).

Finally, let us divide the whole 56-year period into two equal intervals of 28 years each, and find the coefficients of correlation between y and T_6 :

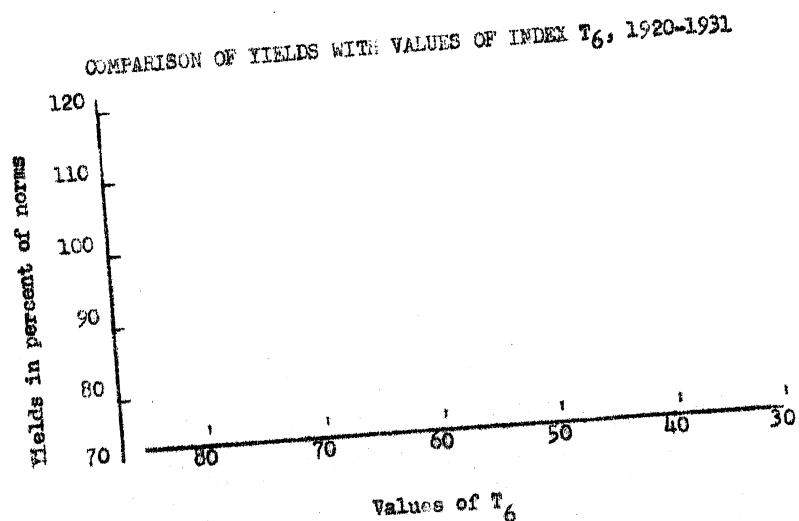
28 years from 1872 to 1899 = -0.677

28 years: 1900-1915 and 1920-1931 = -0.621

All 56 years = -0.639

Thus even the 28-year periods already give fairly stable correlation coefficients, and, a fortiori, the 56-year period. In consequence we are able to maintain that the coefficient of correlation between the yield (y) and the index T_6 , computed for the entire 56-year period and equal to -0.639, not only qualitatively but also quantitatively, closely represents reality.

DIAGRAM 8



90/25

The following table demonstrates the inverse parallelism between y and T_6 .

TABLE 24

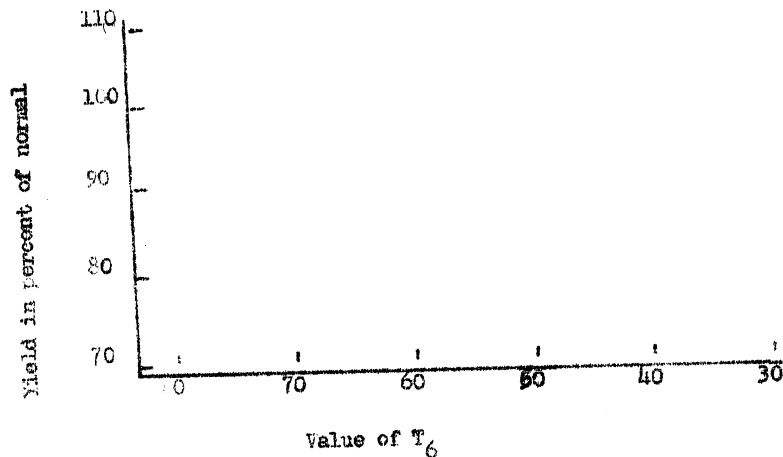
Index T_6	Number of years	Mean yield in percent of normal	Mean value of T_6
75 and over	6	81.3	83.7
65.0 to 74.9	7	87.9	70.3
55.0 to 64.9	18	101.5	59.9
45.0 to 54.9	18	105.8	49.5
under 45	7	108.2	38.3

The higher the numerical value of T_6 , the lower the yield; and as T_6 falls, the yield increases.

Diagram 9, on which the numerical values of T_6 and y have been plotted, vividly portrays this parallelism.

This diagram also indicates something else, namely, it evidences the curvilinear correlation between y and T_6 . The Parallelism noted does not fit within the framework of a linear connection, but requires curves of a higher degree for its precise expression. From the data presented (tables and diagrams) it is clear that a third-degree parabola is the most suitable curve (We note that the curvilinear character of the parallelism, and at that one also represented by a third-degree curve, is also rather clearly delineated on the diagram for the 1920-1931 period.)

DIAGRAM 9

COMPARISON OF THE YIELDS WITH THE VALUES OF T_6 

This equation has the form:

$${}_3Y = 102.276 - 6.646 (t_6 - 6.4) - 1.479 (t_6 - 6.4)^3, R = 0.682. \quad (14)$$

The coefficient of multiple correlation in this case is 0.682. Thus if account be taken of the curvilinearity, a certain small increase in the correlation coefficient results, and consequently also a small improvement in the forecast.

The index T_6 of course, even when improved by curvilinear regression, is inadequate for exact forecasting unless supplemented by other criteria. This is vividly shown by column 5 in table 25, which column contains the values of ${}_3Y$, as computed by the regression equation.

However, in certain cases -- definite cases at that -- its pre-

7

dictions come close to the actual situation. In years when the numerical values of T_6 exceed 65, it may be said with very great probability that favorable spring and summer weather in the producing belt is out of the question and that the weather will bear the definite marks of aridity. In fact, 13 years out of the 56 are characterized by values of T_6 equal to 65 or higher than that, and the crop was not even average in a single one of these years. All of them were under average, though in varying degree. It is likewise very important to note that this is also observed for short periods, regardless of the size of the correlation coefficient that characterizes the period. We have seen that the period 1872-1882 has a correlation of -0.167 between y and T_6 . Nevertheless, in the two years when T_6 exceeded 65 (1876- and 1880), the crops were under the average (95.3 and 87.3). The following period, 1880-1893, is characterized by the high correlation coefficient -0.772. ^{was} T_6 over 65 in four years: 1889, 1890, 1891, 1892; and crops were poor in all of them: 81.8; 92.7; 71.0; 86.8. The third period, 1894 to 1904, likewise with a high correlation coefficient, -0.799, has one year with T_6 over 65, which yielded a harvest sharply under average (85.7).

The fourth period, 1905-1915, has a correlation coefficient of -0.533. Two of its years are marked by a value over 65 for T_6 , and crop yields were definitely poor in both of them (1906 and 1907, with yields of 76.1 and 90.0 respectively).

Finally, in the last, post-revolutionary period, 4 years -- 1920, 1921, 1929 and 1931 -- have T_6 over 65. All these years were under average in varying degree, ranging from 1929, somewhat reduced

to 1921, which was very poor.

In this area the index T_6 may be utilized in practice, as verified on materials of years not comprised in the investigation proper. Such years have proved to be sufficiently effective for forecasts at certain values of T_6 , namely those over 65.

(b) A verification was likewise made for the index B_0 and for its components B_1 and B_{-1} (see Chapter I, pages 52-55, and the note to Table 25).

For the period 1863-1915, as we have seen from the investigation proper, this index has only a low weather forecasting value: the correlation coefficient between y and B_0 is plus 0.355. We have a confirmation for this criterion on a period not included in the investigation proper (1920-1931); the correlation coefficient in this case is hardly different from that given above; it is plus 0.342. The hypothesis is thus confirmed both qualitatively and quantitatively. For the entire 46 years, $r_y, B_0 =$ plus 0.337. Its individual components (B and B_{-1}) correlate as follows with y :

1863-1915	r_y, B	$= +0.263$	r_y, B_{-1}	$= +0.267$
1920-1931	"	$= +0.284$	"	$= +0.334$

	$= + 0.270$	"	$= +0.266$
--	-------------	---	------------

46 years (including the year 1882, which was not comprised in the base period (1863-1915)).

The combination of this criterion with T_6 makes it possible to increase the correlation coefficient somewhat, bringing its value

up to around 0.70 (it may be a little higher or a little lower).

[See Table 25 and notes on following pages]

19. Thus the results of verification, by three methods, of the proposed forecasting system, have proved to be entirely satisfactory. Nevertheless, the author does not consider them decisive. Scientific skepticism allows this system to be treated only as a hypothesis -- a very probable hypothesis -- but still only a hypothesis until the work of verification has been carried to its conclusion.

What still remains to be done in this direction?

Firstly, to verify the third basic index on data of years not comprised in the investigation. This is the index of the difference between the November, January and April precipitation in the Northeast and the Southwest (cf. Chapter IV, pages 91-99). Only after this has been done will it become possible to verify (on data of years not comprised in the investigation) (Due to a shortage of data this was not done) the integral index Θ_{11} which combines all the criteria included in the investigation, and also to verify its resultant forecasting equation: $y = 12.316 + 28.5067 \Theta_{11}$ (of page 99), which has a high correlation coefficient ($R = 0.867$).

Secondly, (and this "secondly" is quite as important as the "firstly") it is necessary to make the transition from the indirect index of the course of the spring and summer weather to the direct meteorological characterizations of that weather. The proposed system has been developed on the basis of the crop data of the pre-

revolutionary epoch, when the rise in productivity took place very slowly and gradually. If even for this period the crop forecasts related only to that part of the dynamics of crop yields that was determined by the meteorological factors, i.e. related only to the fluctuations of the crop yields around the trend line, and could not be given in the form of harvest per desyatina or hectare (cf. Chapter I), it follows that this conventional significance of the forecast becomes more emphasized than ever in our times of socialism and construction, when we are entering an epoch of indubitable, powerful and sharp advances towards higher levels of productivity -- of advances to which the characterization "slow and gradual" is entirely inapplicable. It must likewise not be forgotten that at a certain level of productivity enhancement, the character of the dependence of crop yields on meteorological factors undergoes a fundamental transformation. And so the indices of yields to be awaited, given in this paper in terms of percentage deviations from the trend line, will in the last analysis become their own peculiar type of index of the coming weather in the spring-summer period, inasmuch as they have been established on the basis of data from that epoch when these fluctuations of the yield were still determined by the power of the elemental forces of nature. But if this is so, then would it not be better to employ at once direct indices of the weather to be expected instead of indirect ones? Would it not be better in the future to verify on the materials of years outside the investigation proper, and to deepen, that variant in which the integral index $\oplus$ 1) of the criteria is compared with B_5 (The principal difficulty involved in the small number of meteorological stations during 1871-1882, and also, to some extent in 1916-1924. But the author already disposes of data from 10

stations that operated without a break from 1871 to 1915, though he did not set himself the special task of collecting such material since 1871. It should not in all probability be difficult to increase the number of such stations characterizing the producing belt of the European part of the USSR to 15 or 20, covering the period from 1871 down to the present. It is likewise obvious that the number of criteria to characterize the humidity conditions of the weather should be increased; they should not be confined to rainfall and temperature, but such indices as the number of days with rainfall, the deficit of atmospheric humidity ($E - e$), and the cloudiness, must also be introduced. Unfortunately, such important characteristics as evaporation, per atmometer, duration of sunshine, will hardly be susceptible of inclusion in long time-series, in view of their absence (with rare exceptions) from the data for prior years.) the characteristic of the spring-summer weather? (cf. pages 105-109).

Finally, and in the third place, it is necessary more accurately to define the limits of the territory and of the time over which the forecasting significance of the individual criteria extends.

From a manuscript, completed in 1940.

TABLE 25

Year	y	T ₆	$\frac{T_6 - t_6}{9}$	t ₉ = 6.4	3 ^y	B _c
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1872	98.7	54.1	6.1	-0.3	104.1	
1873	99.7	56.1	6.2	-0.2	103.5	
1874	109.7	57.1	6.3	-0.1	102.9	
1875	84.3	47.0	5.2	-1.2	107.5	
1876	95.3	68.7	7.6	+1.2	92.8	No data
1877	104.6	59.9	6.7	+0.3	100.2	
1878	109.5	33.8	3.8	-2.6	104.4	
1879	91.6	53.3	5.9	-0.5	105.2	
1880	87.3	71.9	8.0	+1.6	89.2	
1881	112.8	64.0	7.1	+0.7	97.0	
1882	102.1	60.0	6.7	+0.3	100.2	29.0
1883	100.5	48.6	5.4	-1.0	107.1	30.5
1884	107.0	41.6	4.6	-1.8	107.5	28.5
1885	89.2	45.0	5.0	-1.4	107.7	26.5
1886	100.0	56.9	6.3	-0.1	102.9	34.0
1887	114.6	48.8	5.4	-1.0	107.1	39.5
1888	109.1	53.3	5.9	-0.5	105.2	32.5
1889	81.8	79.2	8.8	+2.4	81.7	31.0
1890	92.7	83.2	9.2	+2.8	78.8	28.5
1891	71.0	98.7	11.0	+4.6	73.4	27.0
1892	86.8	79.0	8.8	+2.4	82.5	33.0
1893	121.8	50.8	5.6	-0.8	106.5	37.5
1894	122.0	39.1	4.3	-2.1	106.9	36.0
1895	109.8	39.4	4.4	-2.0	106.9	33.0
1896	104.8	55.8	6.2	-0.2	103.5	38.0

TABLE 25 (cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
1897	85.7	71.6	8.0	+1.6	89.2	37.0
1898	98.8	46.1	5.1	-1.3	107.7	30.5
1899	110.0	46.0	5.1	-1.3	107.7	30.5
1900	102.5	59.8	6.6	+0.2	100.9	33.5
1901	86.1	59.7	6.6	+0.2	100.9	36.0
1902	115.3	52.5	5.8	-0.6	105.6	36.5
1903	104.0	48.9	5.4	-1.0	107.1	33.5
1904	118.9	39.7	4.4	-2.0	107.0	34.5
1905	92.6	59.1	6.6	+0.2	100.9	37.5
1906	76.1	77.8	8.6	+2.2	84.1	30.0
1907	90.0	70.3	7.8	-1.4	91.0	29.0
1908	91.5	64.3	7.1	+0.7	97.0	34.5
1909	115.7	49.8	5.5	-0.9	106.8	35.0
1910	106.4	57.8	6.4	0.0	102.3	31.5
1911	81.8	60.3	6.7	+0.9	100.2	28.0
1912	107.9	48.8	5.4	-2.0	107.1	32.0
1913	117.1	54.2	6.0	-0.4	104.7	30.0
1914	88.5	38.1	4.2	-2.2	106.6	21.0
1915	110.9	47.6	5.3	-1.1	107.3	21.5
1920	80.1	84.2	5.4	+3.0	77.5	21.0
1921	71.8	74.9	8.3	+1.9	86.6	19.0
1922	116.1	58.0	6.4	0.0	102.3	21.0
1923	101.4	36.5	4.1	-2.3	105.6	30.5
1924	83.4	62.6	7.0	+0.6	97.8	34.5
1925	115.7	62.0	6.9	+0.5	98.6	31.0

TABLE 5 (cont'd)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
1926	112.2	46.1	5.1	-1.3	107.7	29.5
1927	103.5	49.5	5.5	-0.9	106.8	32.5
1928	106.0	56.8	6.3	-0.1	102.9	35.5
1929	98.7	65.5	7.3	+0.9	95.3	39.0
1930	111.6	59.6	6.6	+0.2	100.9	31.5
1931	85.7	69.4	7.7	+1.3	51.9	27.5

Coefficients of correlation of y (crop yield) with the index T_6 .
 For the same period 1883-1915 (33 years) $r_{y, T_6} = -0.723$

For 11-year periods:

1872-1882 (11 years)	$r_{y, T_6} = -0.167$	1872-1893 (22 years)	$r_{y, T_6} = -0.632$
1883-1893 (11 years)	" = -0.772	1894-1915 (22 years)	" = -0.687
1894-1904 (11 years)	" = -0.779	1920-1931 (12 years)	" = -0.633
1905-1915 (11 years)	" = -0.533	1872-1899 (28 years)	" = -0.677
1920-1931 (12 years)	" = -0.633	1900-1915 (28 years)	" = -0.621

For all 56 years -0.639

The regression equation:

$$3Y = 102.276 - 6.646 (t_9 - 6.4) - 1.479 (t_9 - 6.4)^2 + 0.337 (t_9 - 6.4)^3$$

$$R = 0.682. \quad (14)$$

Footnote to Table 25.

y: yield in percent of the normal yield corresponding to each year.
 T_6 : index composed of observations at Kaxan' station, as indicated
 on page 79.

- a w
- O_1 : temperature of 1st October decade of current agricultural year.
 O_{-1} : temperature of 1st October decade of prior agricultural year.
 J_3 : temperature of 3rd January decade of current agricultural year.
 A_3 : temperature of 3rd April decade of current agricultural year.
 A_{-3} : temperature of 3rd April decade of prior agricultural year.
 a^{FW} : an index characterizing the abruptness of the transition from autumn to winter, as has been pointed out on page 79, the abruptness of this transition is defined by the temperature difference between the last two warm decades and the first two cold ones.

It is entirely clear that the index T_6 is composed of the temperatures of nine decades. The index t_9 , equal to $\frac{T_6}{9}$, is the same index, applied to a single decade. The mean value of t_9 for the 56 years is 6.4. The departures for this mean for each year are shown in column 4. These conversions of the index T_6 into t_9 and $(t_9 - 6.4)$ have been made to facilitate the computation for the third-degree regression equation given at the end of the table:

$$3Y = 102.276 - 6.646 (t_9 - 6.4) - 1.479 (t_9 - 6.4)^2 + 0.337 (t_9 - 6.4)^3$$

Instead of using three-digit numbers and larger (from 33.8 to 98.7) we have to deal in the end, as the result of these conversions, only with two-digit numbers not higher than 4.7. The values of $3Y$ calculated by the regression equation are given in

column 5. They are identical with those that would be obtained if the regression equation had been found directly in terms of T_g .

B_0 is the commencement of unfolding of the leaves at the meteorological station of Bogoroditskoye-Fenino for two years (see pages 52-55, Chapter 1). The index is figured in days from the base date April 1.

THE INFLUENCE OF METEOROLOGICAL FACTORS ON THE FLUCTUATION OF CROP YIELDS IN THE ARID POVOLZH'YE AT VARYING LEVELS OF CULTURAL PRACTICE

CHAPTER I.

WINTER RYE

We begin our exposition by referring to a paper by the famous statistician S. Kharizomenov, printed in Vestnik sel'skogo khozyaystva yugo-vostok (Southeastern Agricultural Herald) in 1911. (The paper is entitled "The influence of rainfall and atmospheric temperature on cereal crop yields in Saratov, Samara and Tambov Guberniyas". Of all the "total" yield for all cereals given in this paper, rye is separately stated only for Saratov Guberniya.) By the use of the statistical classifications, the author established the direct dependence of the rye yields on peasant farms in the former Saratov Guberniya (1881-1905) on the spring rainfall (April, May, June) and on the rainfall in the winter months (November to March inclusive).

Spring Rainfall	Yield in Puds	Winter Rainfall	Yield in Puds
(1)	Per Desyatin (2)	(3)	Per Desyatin (4)
6 years with lowest rainfall (30.4-78.4)	36.4	9 years with lowest rainfall (44.9-78.1)	36.8
9 years with rainfall from 87.1 to 109.8	43.1	9 years with rainfall from 79.6 to 127.2	43.1
8 years with highest rainfall (110 to 227)	53.3	7 years with highest rainfall (130.2 to 169.3)	55.9

[1]	[2]	[3]	[4]
Range for Period	16.9	Range for Period	19.1

The temperature during the winter period showed no influence on the yield, but the temperature conditions during the spring and summer life of winter rye (both for the single months and for the period as a whole) acted in the inverse sense: the higher the temperature, the lower the yield, while on the other hand cool weather favored a good rye crop. For the month of May, Kharizomenov has one reservation to make to the general rule: in years when the May rainfall exceeds 50 millimeters, the influence of higher temperature on the rye crop changes from unfavorable to mildly favorable. His conclusions are very reserved with respect to the significance of autumn rainfall: "The Autumn rains of August, September and October occupy a special position in Saratov Guberniya", we read in his paper, "a minimum autumn rainfall of 70 millimeters is necessary for the successful growth of winter rye. Above this limit it is superfluous, and exerts neither harmful nor appreciably beneficial effect." The table presented by him argues rather for the neutral position of autumn rainfall:

	RYE YIELD
6 years with lowest autumn rainfall (27.8-73.0)	45.8 puds
6 years with autumn rainfall from 88.1 to 103.0	43.3 puds
13 years with highest autumn rainfall (107.6-188.2)	44.0 puds

On generalizing Kharizomenov's deductions, we reach the following conclusions: the spring-summer period was the most im-

[1]	[2]	[3]	[4]
Range for Period 16.9		Range for Period	19.1

The temperature during the winter period showed no influence on the yield, but the temperature conditions during the spring and summer life of winter rye (both for the single months and for the period as a whole) acted in the inverse sense: the higher the temperature, the lower the yield, while on the other hand cool weather favored a good rye crop. For the month of May, Kharisomenov has one reservation to make to the general rule: in years when the May rainfall exceeds 50 millimeters, the influence of higher temperature on the rye crop changes from unfavorable to mildly favorable. His conclusions are very reserved with respect to the significance of autumn rainfall: "The Autumn rains of August, September and October occupy a special position in Saratov Guberniya", we read in his paper, "a minimum autumn rainfall of 70 millimeters is necessary for the successful growth of winter rye. Above this limit it is superfluous, and exerts neither harmful nor appreciably beneficial effect." The table presented by him argues rather for the neutral position of autumn rainfall:

	RYE YIELD
6 years with lowest autumn rainfall (27.8-73.0)	45.8 puds
6 years with autumn rainfall from 88.1 to 103.0	43.3 puds
13 years with highest autumn rainfall (107.6-188.2)	44.0 puds

On generalizing Kharisomenov's deductions, we reach the following conclusions: the spring-summer period was the most im-

portant for the yield of peasant-grown rye (the influence of the rainfall in this period was somewhat greater than that of the winter rainfall, and to this was joined the influence of the temperature), the winter rainfall occupied the second place, while the effect of autumn rainfall was slight or non-existent.

The dependence of the fluctuations in the rye crops of the Povolzh'ye on the meteorological factors was also studied in the published paper of A. M. Protasov "Dependence between the yields of rye and wheat on meteorological conditions in certain years of the Lower Povolzh'ye" in Zhurnal opytnoy agronomii jugo-vostoka, (Journal of Experimental Agronomy of the Southeast), fascicle I, Saratov, 1927; and in my own unpublished paper "Dependence of the fluctuations in the yields of winter rye in former Samara Guberniya on meteorological factors" (This paper was written for volume 3 of the investigation "Crop Failures and the National Economy", published by Ranton; but volume 3 was never published. The work of Protasov and my own work were done independently.)

Both of these papers use a more perfected method, that of correlation.

Protasov's paper relates to the territory of the former Saratov Guberniya. He establishes that the closeness of the connection between the rye yield and the May-June rainfall during 1892-1917 is expressed by the correlation coefficient 0.78.

My own paper gives the following correlation coefficients between the rainfall in the corresponding months and the fluctuations in the rye harvests:

TABLE I

Month	Correlation Coefficients	
	1st Variant	2nd Variant
April	+0.218	+0.214
May	+0.744	+0.755
June	+0.370	+0.387
July	+0.150	+0.196

Note: The rainfall is computed for the first variant from the data of the Monthly bulletins of the Central Geophysical Observatory (GGO). For the second variant, the data in the GGO Yearbook is used instead. The monthly bulletins include a smaller number of stations than the Yearbooks (10-15 stations as against 28-45), but give the data by decades, while the yearbooks give it by months. I took advantage of this point of superiority of the bulletins when I shifted to the characterization of the weather by decades. Comparison of the correlation coefficients computed from both sources was made to verify the suitability of the data from the bulletins. This verification had to be made, since the bulletin data were based on fewer observations than those of the Yearbooks. It will be seen from the comparison that the verification resulted favorably for the Bulletins (we remark that the parallelism between the rainfall indices of the Bulletin and those of the Yearbook is very great for the various years from 1893 to 1915, and is measured by the following correlation coefficients between the two sets of data: 0.985 for the April rainfall, 0.982 for the May rainfall,

0.960 for the July rainfall, 0.929 for the July rainfall. This is an additional proof of the fact that the Bulletins reproduce very satisfactorily the variations in local rainfall from year to year, in spite of the small number of observations they contain.

As we see, the May rainfall had a very strong influence on the rye yields in the former Saratov guberniya: the correlation coefficients by the two variants are almost identical: 0.744 and 0.755. June takes the second place, according to both sources. Finally, April and July, in both variants, are marked by the very feeble influence of their rainfall on the rye yield (0.216 and 0.214 for the April rainfall, and 0.150 and 0.196 for the July rainfall). The aggregate May-June rainfall, correlated with the rye yield in former Samara guberniya, gave a correlation coefficient of 0.78, identical with that for former Saratov guberniya.

Protasov's study also covered other weather indices besides the rainfall: number of days with harmful temperatures (those on which the temperature at 1300 hours was 30 degrees Centigrade or over); and the deficit of atmospheric humidity, measured by the saturation deficit by Dalton's formula ($E-e$), where E is the saturation vapor pressure at the existing current temperature and e is the actually observed vapor pressure (or, what is almost equivalent, the absolute atmospheric humidity). Inclusion of the saturation deficit to supplement the action of the rainfall increased the coefficient of correlation found by Protasova from 0.76 to 0.84 (Protasov does not refer to the question of the significance of the autumn and winter rainfall. But the fact that the spring-summer weather correlates with the yield to the extent of 0.84 is evidence of its unconditionally dominating importance for the

rye yield, for it can be shown mathematically that the influence of all the other factors not correlatively connected with the spring-summer weather indices is expressed by a correlation coefficient no higher than 0.55.) He points out that in processing the data by uyeds, the correlation coefficients have a tendency to increase from northwest to southeast, i.e. with the increasing aridity of the climate.

The fundamental failing of Protasov's study (and of Kharizomenov's as well) consist in his operating with monthly indices. This procedure is rather coarse and makes it impossible to relate the meteorological observations even to the basic stages of plant growth.

In my own paper I did not confine myself to comparison of the rye yields in the former Samara guberniya but passed to a decade analysis of the data. As a result we have the following table of correlation coefficients between the yields of winter rye in former Samara guberniya, 1893-1915, and the rainfall in corresponding decades:

[See table on following page]

Both variants show the predominant influence of the rainfall in the May decades on the rye yield in the former Samara Guberniya, a considerable weaker action of the rainfall in the third decade of April, the June decades and the first July decade, and the almost complete absence of any reaction by the crop to the rainfall of the last two July decades and of the first two in April (The almost identical results obtained with both variants show that it makes no difference which is employed as the basis for the in-

TABLE 2

Periods		1st Variant	2nd Variant
April	1st decade	++0.039	+ 0.017
	2nd decade	+ 0.131	+ 0.133
	3rd decade	+ 0.225	+ 0.237
May	1st decade	+ 0.472	+ 0.440
	2nd decade	+ 0.502	+ 0.499
	3rd decade	+ 0.501	+ 0.525
June	1st decade	+ 0.239	+ 0.189
	2nd decade	+ 0.188	+ 0.140
	3rd decade	+ 0.299	+ 0.267
July	1st decade	+ 0.242	+ 0.231
	2nd decade	+ 0.068	+ 0.024
	3rd decade	+ 0.114	+ 0.073

Note: the 1st Variant represents the correlation of the rye yield with the direct observations in the "Bulletin".

The 2nd Variant is based on the monthly normal rainfall for each year, determined from the more numerous observations given in the Yearbook. These monthly values were then proportionately distributed [by decades] in accordance with the distribution of the rainfall in the bulletin data for the same month and year. We thus obtained a correction to the bulletin indices -- a correlation based on the larger number of observations obtained in the "GGO Yearbook".

vestigation. But inasmuch as the second variant is distinguished from the first by reason of the corrections founded on more numerous data (those of the Yearbook) than the basic tabulation of the observations by decades (the Bulletin data), the investigation therefore uses the second variant.)

This method of working up the observations, decade by decade, in itself allows better orientation in the question as to the growing periods in which the meteorological factors are of special importance. The average date the fields of former Samara guberniya were cleared of snow throughout the 23-year period from 1893 to 1915 was 15 April, the average date of commencement of the sowing of spring cereals was 23 April, of the beginning of the rye blooming period was 16 June, and of the commencement of the harvest was 18 July (I give these dates from data in the agricultural surveys issued by the former Department of Agriculture under the title "--- Year in Agriculture" and the data of the current agricultural statistics of the Guberniya Zemstvo of Samara Guberniya.)

I have no systematically arranged data available on the commencement of the heading of rye, but the beginning of this phase can readily be determined from the beginning of blooming. Since rye begins to form heads about 10 days before blooming, we may fix the middle of the first June decade as the beginning of heading of winter rye in the former Samara guberniya during the 1893-1915 period (cf. Appendix 1).

Thus the favorable influence of the rainfall on the yield of winter rye in the former Samara guberniya commenced at the moment that the sowing of the early spring cereals began, i.e. as soon as the temperature of the air had become high enough (We regard the date of commencement of sowing the early spring cereals as an in-

direct index of the beginning of the energetic vegetation on rye. Of course, if we had direct indications of this beginning (for instance, by noting the time it starts to turn green, or the time it starts to grow tall, we would not need this indirect index, but I had no such data at my disposal for former Samara guberniya.) and ended at the arrival of the maturation phase, about ten days before the commencement of harvesting. The most considerable positive influence of rainfall falls in the heading period: the three May decades have the highest correlation coefficients.

The comparison of the yields with the rainfall by decades thus contributes an improvement to the investigation, permitting more accurate orientation in the details of the dependence of the yield on meteorological factor. However results of an approximate character may be given not only by comparison of the yields with the monthly indices of the meteorological elements, but even with fixed calendar decades. Neither of such comparisons may satisfy the requirements of having as great a degree of accuracy in and details of, as possible in the investigation.

For instance, if we analyze the influence of the third June decade on the yield of rye, we are using the rainfall during different vegetation periods in the different years as a basis of comparison with the yield. For 1906 the third June decade was the period of maturation, in 1901, it was the period of ripening, but in 1895 and 1904, the rye had hardly headed fully by the beginning of the third June decade, and it only commenced to bloom in the middle of that decade. As we have seen, all three May decades are outstanding in their influence on the rye crop. Since rye begins on

the average to bloom on 16 June in Samara guberniya, and it begins to head on the 9th to 12th day before beginning to bloom, it is clear that the rainfall before the beginning of heading must be of the greatest importance for the rye crop in Samara guberniya. This general conclusion, however, cannot be further detailed, if we confine ourselves to a fixed calendar periods. Similar correlation coefficients for all three May decades may be obtained for one of two reasons: either because all of the 10 day rainfall periods in May before heading begins are of equal importance, or as a result of the shifting of a shorter but critical vegetation period into the May decades. In some years this vegetation period comes in the first decade of May, in other years in the middle decade, and, finally, in years when the development of rye is delayed it comes in the last May decade. Thus when we compute the correlation coefficients for the May decades over a period of 23 years, we may obtain almost equal values for these coefficients even in case this short vegetation period is shifted.

From the data of the census statistics and the materials of the agricultural surveys of the former department of agriculture, I determined for each year the dates of the basic periods of vegetation: (1) The date of emergence of the winter rye from under the snow; (2) The date when early spring sowing commenced; (3) the beginning of head formation on the rye, and (4) the beginning of its harvesting (see appendix 1, point 3a).

From these basic periods we can then pass to the short periods and we have tabulated the meteorological observations in accordance with these periods.

The period from the emergence of the winter crop from beneath the snow until the beginning of the sowing of the spring crops in the period of spring germination of the winter crop; the period from the beginning of the sowing of the spring crops until the commencement of blooming is the period of the energetic vegetation of rye and the trunk formation; the period from the beginning of blooming until the harvest is the period of bearing, the period of kernel formation of the grains.

I next fixed nine shorter vegetation periods as follows:

- 1st period: Early spring growth of rye.
- 2nd period: Energetic vegetation of rye begins.
- 3rd period: Energetic vegetation of rye continues.
- 4th period: Shooting begins.
- 5th period: Shooting continues.
- 6th Period: Heading.
- 7th Period: Blooming.
- 9th Period: Maturation.

The details of the division of the short periods are given in appendix 2. How the meteorological 10-day indices were applied to this period is also described there. Here I shall only present the results of the comparison of the rye crops with the rainfall falling during these periods. The correlation coefficients between the yields of rye and the rainfall of each period are presented in table 3.

TABLE 3

PERIODS	<u>CORRELATION COEFFICIENT</u>		
	<u>RAINFALL</u>		<u>TEMPERATURE</u>
	1st Variant [2]	2nd Variant [3]	2nd Variant [4]
[1]			
1st period of early spring growth	+0.046	+0.023	+0.393
2nd period) of energetic	+0.213	+0.184	+0.197
3rd period) vegetation	+0.540	+0.536	-0.281
4th period) of shooting	+0.787	+0.810	-0.526
5th period)	+0.457	+0.553	-0.582
6th period of heading	+0.258	+0.257	-0.095
7th period of blooming	+0.448	+0.355	-0.410
8th period of ripening	+0.227	+0.187	-0.419
9th period of maturation	+0.009	+0.041	-0.335

In the first variant the rainfall was compared with the indices of yield as given in the statistical materials. In the second variant the influence condition of the sowings at the emergence from the snow (as was done -- see Appendix I) was excluded from the data of the yield. The second variant thus gives us a better opportunity to follow the action of the spring and summer weather. Both variants agree very closely among themselves.

This is explained by the fact that the spring and summer

weather in the arid Povolzh'ye in general, and on the territory of the former Samara guberniya in particular, has a dominating influence on the formation of the rye harvest, as compared with the conditions of the autumn and winterweather. For both variants the rainfall during the period of incipient shooting exerts an outstanding and very considerable positive influence on the rye crop. Taken together with the period before it and the period following it, the shooting period composes the month-period with the highest correlation coefficient.

Table 3 also gives the correlation coefficient between the harvest of winter rye and the temperature conditions (following the second variant). Appendix 4 gives details of how the temperatures were recorded.

Finally, table 4 gives the partial correlation coefficient between the yield of rye and the rainfall, eliminating the influence of the temperature, on the one hand; and between the yield of rye and the temperature conditions, eliminating the influence of the rainfall on the other hand.

Thus the influence of each factor may be traced in its pure form, uncomplicated by correlation with the other factor.

[See Table 4 on following page]

As will be seen from this table the positive effect of rainfall strongly increases and reaches a maximum in the 4th period, about 40 or 40 days after the emergence of the winter rye from under the snow. The positive influence of rainfall then declines and breaks off at the period of ripening and maturation (when the

TABLE 4

<u>PERIODS</u>	<u>PARTIAL COEFFICIENTS OF CORRELATION (2nd Variant)</u>	
	<u>RAINFALL</u>	<u>TEMPERATURE</u>
[1]	[2]	[3]
1st period	+0.083	+0.400
2nd period	+0.268	+0.276
3rd period	+0.484	-0.098
4th period	+0.762	-0.341
5th period	+0.312	-0.375
6th period (heading)	+0.243	+0.640
7th period (blooming)	+0.163	-0.271
8th period (ripening)	-0.035	-0.383
9th period (maturation)	-0.036	-0.334

correlation coefficients are hardly different from zero, and have the minus sign at that.) In the first period (that of the spring germination of winter rye) the temperature has a positive action, which is already weaker during the second period, and, commencing with the third period, all of the increasingly harmful influence of high temperatures is revealed. During the period of head formation, when the formation of the plant body has already been completed, the influence of the temperature on the rye crop falls sharply, and is again activated on the arrival of the second basic period, that of kernel formation. During the period of the formation of the body, the influence of rainfall on the rye crop in the former Samara guberniya was considerably stronger than that of the temperature, but during the bearing period the action of the temperature conditions on the yield advances the first place.

We put the question how the action of the rainfall and the temperature during the time of the spring and summer life of the rye plant is in general reflected in the yield of rye. The question is not solved by a simple summation of the meteorological elements from the moment the fields are free from snow down to time of the harvest, since their importance during the several periods of plant development is not at all uniform. The method of least squares leads us to the establishment of a regression equation in which the yield is considered as a function of the variables that influence its size.

This equation could make it possible for us to take into consideration the value of each factor individually and of all of them taken together. However, in our investigation, this method in its

pure form is inapplicable, since it would be necessary to set up a regression equation in 18 variables, which would be, from the technical point of view, a very laborious task, but also mainly, because, theoretically, it is rather more than doubtful that such an equation, with 18 variables and 23 observations, would be at all admissible. For this reason we are compelled to have recourse to another method, namely that of combining the fractional vegetative periods into a larger one, thereby reducing the number of variables, and then applying the method of least squares. Based on the combination of short vegetative periods, we thus build up a system of "weighted" meteorological indices, based on the coefficients of partial correlation. It is indubitable that the significance of the factors for the yield is some function of these coefficients. The approximate function may be sought empirically. In the end, the system of proportional weighting by the squares of the partial correlation coefficients proved more suitable (for details see appendix 5). As a result we obtained 3 variables: (1) a generalized index of the rainfall, designated by the symbol P_s [Precipitation Vsesenny - spring rainfall] (2) a generalized index of the temperatures of these decades in which the temperature affects the yield adversely, designated by T_s , (3) the index t_x for the first two decades during which the influence of the temperature on the yield of rye is positive. The regression equation that connects the yield with these meteorological indices, is as follows:

$$q = 1.04191 + 0.1893P_s - 0.7034 T_s + 0.0850 t_x; R = 0.935 \quad (1)$$

The coefficient of multiple correlation of these three fac-

tons reaches a very high value, namely 0.935. The symbol q denotes the yield of winter rye after eliminating the influence of the condition of the sowing at emergence from the snow.

The inclusion of forth variable, the influence of "haze", "pom-
okh", dry mist (I determined the influence of the haze from the data of the current ~~genstvo~~ statistics of the former Samara guberniya, together with the above cited agricultural surveys of the former Department of Agriculture. Those years when the rye did not suffer at all from haze were denoted by 3 balls, while 2.75 balls denoted a slight damage from haze, 2.50 balls an appreciable damage, and 2.00 balls a serious damage. These characteristics are given in further detail in table 6.) leads to a further insignificant increase in the correlation coefficient to 0.942:

$$q_1 = 7,6031 + 0.1880 P_s - 0.6592 T_s + 0.0854 t_s + 0.7434 R; R = 0.942$$

(see appendix 3, in which a summary of all the basic meteorological elements is given).

The generalized index of rainfall during the spring and summer period, P_s , correlates with the rye crop y , from which the influence of the state of the sowings at emergence from beneath the snow has not been excluded, to the extent of plus 0.839, and its correlation with the temperature T_s is 0.779.

The combined influence of rainfall (P_s) and temperature (T_s) on the crop is characterized by the coefficient of correlation 0.924. This evidences the tremendous importance that the weather of the spring and summer period had on the yield of rye in the former Samara guberniya. However as we shall see below, the autumn weather

also left its mark on the results of the rye harvest.

The data of the zemstvo statistics and of the agricultural surveys of the former Department of Agriculture allow us to fix the time when the sowing of the winter cereal crops began as well as the condition of the winter sowings before their emergence from beneath the snow (see appendix 1). The overwhelming importance of the spring and summer period for the harvest of winter rye compelled us to put the question as to the radical exclusion of the influence of the spring and summer weather in an analysis of the autumn period. Unless we did this, any accidental coincidence of the autumn indices during short vegetation periods with the increased or diminished indices of spring and summer weather that increase or diminish the yield, might distort the picture of the relative importance of the meteorological elements during the short vegetation periods in the autumn.

The relative importance of the latter was determined by their comparison not with yield itself (on the results of which the spring and summer weather exercises the dominating influence), but with the condition of the winter crops in late autumn, which of course does not at all depend on the meteorological conditions of the following spring.

The correlation coefficients between the autumn rainfall and autumn temperature conditions, on the one hand, and the condition of the spring sowings before their emergence from the snow, on the other hand, have the following values:

[See Table 5 on following page]

TABLE 5

PERIODS	<u>Rainfall and win-</u>	<u>Temperature and</u>	<u>PARTIAL CORRELATION COEFFICIENTS</u>	
	<u>ter crop condition</u>	<u>winter crop con-</u>	<u>Rainfall and win-</u>	<u>Temperature and</u>
		<u>ditions</u>	<u>ter crop condition</u>	<u>winter crop con-</u>
				<u>dition</u>
[1]	[2]	[3]	[4]	[5]
Decade before beginning of sowing	+0.112	+0.054	+0.101	-0.024
1st decade after beginning of sowing	+0.678	-0.233	+0.663	+0.139
2nd decade after beginning of sowing	+0.453	-0.094	+0.501	+0.256
3rd decade after beginning of sowing	+0.326	-0.298	+0.479	+0.463
4th decade after beginning of sowing	-0.001	+0.591	+0.411	+0.678
5th and concluding decade	+0.142	+0.342	+0.291	+0.418

The coefficient of partial correlation show that the rainfall, the importance of which increases sharply from the beginning of the sowing, has a positive importance for the condition of the sowings in late autumn, but that this positive effect declines regularly as autumn advances, while the positive influence of the temperature on the condition of the winter sowings increases and reaches a maximum value during the fourth decade from the beginning of the sowing. These dependences completely agree with the observations of farmers, who consider that a sufficient amount of rainfall especially during the time of the sowing, and warm weather during the tillering phase, after the beginning of sowing, are conditions for the satisfactory state of the wintersowing up to the end of autumn. The reduction in the influence of the rainfall and the increase in the positive significance of the temperature as the interval from the beginning of sowing increases, is completely explained by the fall in temperature as the autumn advances. Essentially, the 40 day period after the beginning of sowing of the winter crops determined the condition of these crops in the former Samara area until their emergence from the snow.

In order to find the generalized index for the rainfall (P_A) [written either P_0, T_0 or P_0, T_0 (zero) in Russian (Ocenny, Autumn)] we take the rainfall of these 5 decades with different weights. The weights are fixed proportionately to the square of the coefficients of partial correlation in just the same way as was done to find the generalized index of precipitation for the spring and summer period. We proceed in analogous manner to find the generalized index of temperature conditions (T_A). The correlation coefficient between the condition of the winter sowings in late

autumn and the generalized index of autumn rainfall (P_a) is + 0.595, while the coefficient of correlation between the condition of the winter sowings and the temperature conditions is +0.465. The partial correlation coefficients between: the rainfall and the condition of the winter sowings, after eliminating the influence of the temperature, is +0.861; between the temperature and condition of the winter sowings, eliminating the influence of the rainfall, it is +0.829. The coefficient of multiple correlation between the combined influence of both factors and the condition of the winter sowings before their emergence from the snow is 0.887.

In conclusion we give a synthesis of all the results obtained above. Table 6 contains the "generalized" indices of rainfall and temperature for the autumn period and the spring-summer period, and then, using the method of least squares, we determine the yield of rye as a function of all these factors.

The meaning of the generalized indices will now be recalled. P_s is the index of rainfall during the spring-summer period, T_s is the index of the "negative" temperature conditions for the same period, t_s is the index of those weakly positive conditions that occur in early spring, and M is the index of "hazes". For the autumn period we have two indices: P_a characterizes the amount of rainfall, while T_a characterizes the temperature conditions. The system of "generalized" indices has been set forth above. It is based on the principle of eliminating the influence of the autumn and winter weather on the spring-summer period, since this weather is already reflected in the condition of the winter sowings on emergence from the snow, and

excluding the action of the spring-summer weather from the investigation of the influence of the autumn weather.

[See table on following page]

The regression equation in which the yield (y) is regarded as a function of the six above named factors we have mentioned, is as follows:

$$Y_1 = 13.6284 + 0.1700 P_s - 0.7908 T_s + 0.1335 P_a + 0.6362 \\ M = 0.0672 t_s + 0.0648 T_a. \quad (III)$$

The coefficient of multiple correlation between the yield and the combined action of the six meteorological factors reaches the very high value of 0.955 (see diagram 1).

When we examine this regression equation that completes our investigation, we see: (1) that a unit of rainfall in the autumn is of less importance for the yield than a unit of rainfall during the spring summer; (2) that the effect of the rainfall during the latter period is supplemented by the very respectable influence of the temperature, acting in the same direction upon the yield as the rainfall does, namely the negative effect of drought is usually intensified by high temperatures and, on the other hand, the favorable influence of abundant rainfall is supplemented by the action of moderate temperatures; to this there is added during the summer period the influence of the haze as well; (3) that the effect of the rainfall during the autumn is supplemented to a very slight extent by the positive influence of the temperature, to which likewise corresponds a similarly weak posi-

TABLE 6

Year	Yield of rye in tseentrer per hectare	P_a	T_a	P_s	T_s	t_s	H	y_1	$y - y_1$
1	2	3	4	5	6	7	8	9	10
1893	5.26	10.6	12.8	10.9	17.8	3.3	2.50	5.47	-0.21
1894	5.76	10.4	14.8	9.8	18.3	8.2	3.00	5.63	0.13
1895	6.07	12.8	9.9	13.4	18.0	5.6	3.00	6.31	-0.24
1896	6.36	13.8	13.4	22.4	18.9	7.2	2.50	9.28	-0.92
1897	4.04	8.1	13.5	9.2	20.3	5.7	3.00	3.39	0.65
1898	1.81	10.0	13.0	5.6	21.8	6.4	3.00	1.94	-0.13
1899	6.62	17.5	9.7	11.1	18.2	8.2	2.50	6.23	0.39
1900	7.01	13.8	14.7	16.1	18.7	7.0	2.00	6.11	0.94
1901	3.54	14.7	11.8	3.8	20.4	7.0	3.00	3.25	0.29
1902	3.86	10.1	9.4	10.7	20.2	5.8	2.50	3.41	0.45
1903	7.78	17.5	7.3	24.4	19.8	8.9	2.75	7.28	0.50

TABLE 6 (Cont'd.)

1	2	3	4	5	6	7	8	9	10
1904	7.15	8.7	8.1	13.3	17.1	0.3	2.15	6.49	0.66
1905	2.51	3.7	11.0	7.0	19.8	7.9	3.00	2.81	-0.30
1906	0.94	13.8	14.6	1.9	21.1	5.2	2.00	1.68	-0.74
1907	4.78	16.8	10.2	12.8	19.9	7.1	2.50	5.04	-0.26
1908	4.57	10.8	10.5	13.3	18.5	0.4	2.50	5.67	-1.10
1909	7.66	8.9	12.5	24.8	18.3	8.3	3.00	7.84	-0.18
1910	4.79	1.6	13.1	15.0	19.1	9.9	2.50	4.39	0.40
1911	1.57	12.4	11.9	5.2	22.0	8.2	2.00	1.21	0.36
1912	7.34	16.0	10.4	16.9	19.0	5.8	2.75	6.42	0.92
1913	7.12	8.1	12.8	22.5	18.3	8.3	2.75	7.20	-0.08
1914	4.20	13.5	8.7	15.3	19.1	7.5	2.75	5.74	-1.54
1915	8.00	18.2	11.4	17.8	18.1	7.9	3.00	7.95	0.05

tive influence of the temperature in early spring.

However, we can only get a complete idea of the influence of the several factors on the yield of rye if we take into consideration the extent of the influence each of these factors exerts, according to the sum of those modifications introduced by them into the yield during the course of the period under study.

Indeed we see that a unit of rainfall during the spring-summer period has about 25 to 30 percent more importance than during the autumn. (The coefficient in the regression equation with P_a is equal to 0.1325, whereas with P_s it is + 0.1700). This ratio may be increased or decreased according to the amount of fluctuation in the series P_s (the index of spring and summer rainfall) and P_a the index of autumn rainfall). If the dispersion of the values of the spring-summer rainfall is greater than the dispersion of the values of the autumn rainfall, then the strength of the influence of the spring-summer rainfall is augmented with respect to the action of the autumn rainfall, and in the reverse case, that is, when the contrary relation obtains between the dispersions of P_s and P_a , the effect of the spring-summer rainfall is weakened. A wider range of fluctuation of the meteorological factors also causes greater variation in the yield, regardless of what the coefficients of the variables in the regression equation may be. The full effect of any factor is a function on the one hand of the size of this coefficient, but, on the other hand, it is also a function of the variability of the factor itself. For a final determination of the force and significance of each of the meteorological elements included in this investigation we put the regression equation into a somewhat different form, namely we express the variables in terms

of their deviations from their corresponding means.

$$y_1 = +0.1700(P_s - 13.17) - 0.7908(T_s - 19.26) + 0.1335(P_s - 11.85) \\ + 0.6362(M - 2.66) + 0.0672(t_s - 7.40) + 0.0648(T_s - 11.54). \text{ (IV)}$$

The extent of the influence of each factor on the yield in each year is determined by multiplying the corresponding coefficient in the regression equation by the deviation of the factor from its mean value. Thus for instance, in 1906, the index of spring-summer rainfall was 1.9, consequently its deviation in that year from the mean value of the rainfall for the entire 23 year period (13.2) is equal to 11.3. The coefficient of P_s is 0.1700. Therefore the reduction in the yield as compared with the average yield over a period of years (5.16 Tsentner per Hectare), in connection with the poor rainfall in 1906, is equal to $0.1700 \times 11.3 = -1.92$ tsentner; and in 1903, the index of spring-summer rainfall is expressed by the value 24.4 millimeters. Consequently the increase of the 1903 yield over the mean level of the yield, caused by the good moisture conditions in the spring, is equal to $0.1700 \times 11.2 = +1.90$ tsentner.

These calculations were made on the rainfall for each year. The sum total of the change in the yield of rye introduced by the spring-summer rainfall is equal to 19.30 tsentner (This total may be calculated by the formula $S = 0.1700 \sum |P_s - 13.2|$, where $\sum |P_s - 13.2|$ represents the sum of the absolute values of the deviations of P_s from its mean 13.2. Thus sum may be computed without having to find these deviations separately for each year by using the formula:

$$\sum |x - M_0| = nM_0 - (x_1 + x_2 + x_3 + \dots + x_n) + (x_{n+1} + 1 + x_{n+2} + \dots + x_n) - (n-m) \text{ where } n \text{ is the number of the observations, } M_0 \text{ the}$$

arithmetic mean of the values of x and $(x_1 + x_2 + x_3 + \dots + x_n)$ is the sum of those values of x that are less than M_0 . The latter formula may be simplified, for the sum of the deviations below the arithmetic mean is equal to the sum of the deviations above it.

Of all the methods proposed, the method of determination by the expanded formula is most precise in the calculation of the decimals. We also use this formula in cases where it is required to find the entire amount of influence exerted by any given factor during the whole 23 year period. In such cases, when we must know the action of such factor in each separate year, these computations should be made directly. Examples of such calculations are presented above.), which is equivalent to an average of 0.84 tsentner per year. In Table 7 we give the quantitative characteristics of the influence of these factors on the yield of rye separately for each year for P_s and T_s and also give their combined effect (P_s and T_s).

[See Table 7 on following page]

Such generalized indices, computed for all six factors, are as follows:

METEOROLOGICAL ELEMENTS

(1)

SIZE OF CHANGES CAUSED IN YIELD IN TSENTNER PER HECTARE (AVERAGE PER YEAR)

(2)

TABLE 7

INFLUENCE OF THE SEPARATE FACTORS ON THE YIELD, EXPRESSED IN TSEMTERS PER HECTARE

Year	P _s	T _s	P _s + T _s	Year	P _s	T _s	P _s + T _s
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1893	-0.39	+1.18	+0.79	1905	-0.88	-0.39	-1.27
1894	-0.57	+0.79	+0.22	1906	-1.92	-1.42	-3.34
1895	+0.03	+1.02	+1.05	1907	-0.06	-0.47	-0.53
1896	+1.56	+0.21	+1.87	1908	+0.01	+0.63	+0.64
1897	-0.68	-0.79	-1.47	1909	+1.97	+0.79	+2.76
1898	-1.32	-1.97	-3.29	1910	+0.30	+0.15	+0.45
1899	-0.35	+0.88	+0.53	1911	-1.36	-2.29	-3.65
1900	+0.49	+0.47	+0.96	1912	+0.62	+0.23	+0.85
1901	-1.59	-0.86	-2.45	1913	+1.58	+0.79	+2.37
1902	-0.42	-0.79	-1.21	1914	+0.35	+0.15	+0.50
1903	+1.90	-0.39	+1.51	1915	+0.78	+0.94	+1.72
1904	+0.01	+1.73	+1.74				

(1)

(2)

Rainfall of spring-summer period	0.84
Temperature of same period	-0.83
Rainfall of autumn period	0.45
Influence of haze	-0.15 (as already pointed out, years unaffected by haze

are designated by us with 3.00 balls. As the influence of haze increases, we reduce the balls successively to 2.75, 2.50 and 2.00. The positive coefficient of M in the equation therefore indicates that the better the yield, the smaller the influence of haze. [The yield improves as the influence of haze declines.] To eliminate misunderstandings in the final result, we changed the plus sign to a minus sign. The same result would be obtained if instead we had designated by the highest balls the years with highest harmful influence of haze, and reduced the balls as the harmful influence declines.)

Temperature of autumn period	0.11
Temperature of early spring period	0.09

The spring-summer rainfall and temperature conditions exerted the highest influence on the yield of rye (+0.84 and -0.83). The negative effect of the latter is equal to the positive effect of the former. The exceptionally dominant position of the spring-summer weather is thus confirmed by the analysis completing the investigation. The combined influence of precipitation and temperature during the spring-summer period is expressed by the index 1.52 tseentner per hectare (Column 4 of table 7, designated by the symbol $P_s + T_s$, was added up, and the total then divided by 23, the number of years

covered by the study. The index obtained -- 1.52 -- differs only very slightly from a simple summation of the effects of the rainfall and temperature ($0.63 + 0.84 = 1.67$) since the signs P_s and T_s in Table 7 are the same in the vast majority of cases.) The regression equation in which the yield is made dependent only on these two factors is as follows:

$$y = 17.5403 + 0.1880 P_s - 0.7724 T_s$$

The coefficient of multiple correlation for the joint action of these factors [on the yield] is 0.924.

The autumn rainfall occupies third place. Its effect is considerably less than that of the spring-summer rainfall (0.45 as against 0.84), and is more than three times weaker than the combined effect of rainfall and temperature during the spring and summer. The regression equation that combines the effect of these three factors is as follows:

$$y = 17.3195 + 0.1746 P_s - 0.8190 T_s + 0.1109 P_a$$

The coefficient of multiple correlation is increased to 0.945; the partial coefficient of correlation for the autumn rainfall is +0.52. The remaining three factors have only a very slight influence on the yield of rye: even taken together, in the aggregate, the index of their total effect during the entire 23-year period is less than that of the autumn rainfall ($0.15 + 0.11 + 0.09 = 0.35$, as against 0.45). The weak positive effect of the autumn temperature conditions on the final yield is striking, in view of the undoubted importance of heat for the growth and development

of winter crops and for their condition during the autumn period (see page 129). This marked enfeeblement of the positive effect of autumn warmth on the final yield of rye becomes entirely understandable if we consider the influence of autumn temperature on the soil moisture stocks. During the autumn, stocks of moisture are accumulated in the deep horizons of the soil. The size of these stocks is important for the spring development of rye. Increased autumn temperature, on the one hand, favors tillering and, in general, the growth and development of the winter sowings during the autumn months, but on the other hand, also acts unfavorably on their future life by reducing the soil moisture stocks. Autumn rains, however, do not suffer from this contradiction and act only in one direction: they exert a favorable influence both on the condition of the winter crops during the autumn and on the soil moisture stocks (see Appendix 6 for supplementary explanations).

Analysis of the dependence of the fluctuations of the rye crops on the meteorological factors, in the former Samara guberniya, in the pre-revolutionary period, -- an analysis based on the principal vegetative periods and then broken down in detail by short time-intervals-- has thus brought us to the following conclusions:

(1) The basic factors that determined the fluctuations in the yield of rye were the spring-summer weather conditions: rainfall and temperature. Abundant rainfall and moderate temperatures during this part of the year caused upward deviations in the crop, while drought and high temperatures caused crop failure. (2) Autumn rainfall played a secondary but still quite appreciable role. The influence of the remaining factors was very slight. (3) The spring-summer period is characterized by a rise in the influence of meteorological

factors attaining maximum intensity during the period when shooting begins. This influence then sags considerably up to the period of spikelet formation, or heading, and then, with the arrival of the second basic period, that of kernel formation, it undergoes a new wave of rise and fall. (4) The weather during the period of body formation (until blooming) is more important for the yield than that during the bearing period. Rainfall is the dominant influence during the former period; temperature during the latter.

The research herein set forth was completed by me in 1929. In the following years I carried out a verification of its conclusions with respect to the spring-summer period, bearing in mind the fact that those conclusions had the following weak points: (1) since I was using ready-made totals by decades, I was compelled to have recourse to artificial apportionment of the data from parts of those decades in studying the shorter vegetative periods (see Appendix 2); (2) the project covered the entire territory of the former Samara guberniya, while this area is divided, according to productivity and meteorological factors, into the more arid South (Nikolayevsk and Nova-Uzensk uyezds) and the North, which is characterized by a milder climate (the remaining 5 uyezds); and (3) the method of least squares was employed "not according to the strict rules of the art".

In proceeding to the verification of the conclusions, I had the daily weather records of the Polibino station (North) and M. Uzen' (South) together with those of a number of other stations either on the border of Samara guberniya or very close to it. These other stations were weighted 0.5 or 0.25 by comparison with the Polibino and M. Uzen' stations. I was thus delivered from the "dismemberment" of the ten-day meteorological indices and was enabled to compute the

indices independently for the short vegetation periods.

All of the conclusions presented above, and more particularly the wavelike character of the variation in the effectiveness of the spring-summer meteorological factors, were also confirmed by this more refined utilization of the observations of the meteorological stations. The correlation coefficients obtained in this case were somewhat higher.

We present two resultant equations set up by the method of least squares, in which the degree of influence on the yield exerted by moisture conditions during the several vegetation periods is traced (first equation) and by the saturation deficit (second equation).

The first equation is as follows:

$$2 = 5.976649 + 0.245166 P_w + 0.150562 h_2 + 0.298906 h_3 \\ + 1.300376 h_4 + 0.277016 h_5 + 0.018568 h_6 + 0.503985 p_{h7} \\ + 0.305942 p_{h8} - 0.203303 m_{h9}; R = 0.956$$

In this equation the fluctuations q in the yield of rye (from which the influence of the condition of the winter sowings at emergence from the snow has been eliminated) are placed in dependence on nine factors: (1) the winter rainfall, combined with that in late autumn (the decades with temperature below 5 degrees) and in early spring (the rainfall of the first period, that of the spring germination of rye): P_w ; (2) h_2 ; the rainfalls during the second period -- the beginning of the energetic vegetation of rye; (3) h_3 ; the rainfall during the third period; (4) and (5)

h and h : the rainfalls during the period of shooting; (6) h_b , the rainfall [UVLAZHNENIYE, literally "moistening". While this might conceivably include the effect of the soil moisture stocks, careful reading of the context and of the preceding text covering the study proper will convince one that this term is used here as a simple synonym for OSADKI.] during the heading period; (7), (8) and (9): ph_7 : the rainfalls during the blooming period; ph_8 : the rainfall during the ripening period; ph_9 : the rainfall during the maturation period.

The coefficients of the variables in this equation are given in such a form that their values serve to indicate the relative influence of each factor on the fluctuations of the rye crop, and consequently we may determine the "weight" for each factor from these coefficients:

Factors	Coefficients in equation	"Weights"
P	0.254166	0.0742
h	0.150562	0.0456
h	0.298906	0.0905
h	1.300376	0.3936
h	0.277016	0.0839
h	0.018568	0.0056
h	0.503.985	0.1525
h	0.305942	0.0926
h	-0.203303	-0.0615
Total	3.303824	1.0000

This regression equation, as has been indicated, is set up

strictly in accordance with the rules of the method of least squares, without those simplifications that were made in the original work, and, consequently, takes the mutual correlation between the factors into account. As we see, this more refined version confirms the basic conclusions of the original work. The importance of rainfall first increases to reach its maximum during the fourth period (the early shooting phase) and then declines, falling to minimum during the period of the completion of plant-body (TELO) and spike formation (the heading period), thence once more becoming activated with the arrival of the grain formation period, subsequently dropping off again during the ripening phase, and finally, with the arrival of maturation, being characterized by a mildly negative action on the yield of rye (a detail that was passed over in the original working up). As we see, each of these periods, standing alone, is marked by only an insignificant effect on the final result of the crop, except for the fourth period (h_4), but even this period does not exercise a decisive influence on the yield.

The rainfall of this period takes part to the approximate extent of 40 percent [See my work "A contribution to the question of forecasting the spring and summer weather", Introduction, Pages 50-52). Its application in this instance reduces the coefficient of correlation for the aggregate action of the factors included in the study to 0.929. But even so it still remains very high.

Let us see what result will be obtained by working up the data on the saturation deficit. The formula for the saturation deficit is $E - e$, where E is the saturation vapor pressure at the given

actual temperature, and e is the observed vapor pressure at the same temperature (or, what is almost the same thing, the absolute humidity of the atmosphere). The first member of this formula is a function of the temperature. Consequently the formula also takes account simultaneously of both temperature conditions and atmospheric humidity and therefore better defines the atmospheric conditions of plant existence than does the temperature by itself. The saturation deficit is taken not at noon but at 1300 hours, since it expresses the degree of atmospheric dryness better at this time. The regression equation in which the yield is taken as a function of the saturation deficit during the several vegetative periods, commencing with the third, is as follows:

$$\begin{aligned} q = & 9.49726 - 0.04747 (E - e)_3 - 0.84933 (E - e)_4 - 0.90028 (E - e)_5 \\ & - 0.11564 (E - e)_6 - 0.11706 (E - e)_7 - 0.54725 (E - e)_8 - \\ & - 0.11915 (E - e)_9 - R = 0.918. \end{aligned}$$

As we see, there is a minus sign before all the variables in this regression equation, which indicates that there is an inverse dependence of the yield on the saturation deficit, but the extent of this dependence is subject to extreme variation: it is negligible in the third period, sharply rises in the fourth, rises even more in the fifth, then falls in the sixth and seventh periods (heading and blooming), revives in the eighth period (ripening) and finally drops in the time of maturation. Thus this work-up confirms the conclusions of the last one: we note the wavelike character of the value of the indices of atmospheric aridity, but in addition this new variant, worked out according to all the rules of the method of least squares, introduces a few interesting nuances. We note that the maximum values of the saturation de-

ficit lag one period behind those of the rainfall.

During the period of plant-body formation, the rainfall has its highest positive effect in the fourth period (the first half, from early shooting to the beginning of heading), while the maximum negative influence of the saturation deficit falls in the fifth period (during the period after the heading). During the grain formation period, the rainfall is most important in the seventh period (that of blooming), while the corresponding maximum negative effect of the saturation deficit falls in the following eighth period (that of ripening). The distribution of the corresponding maxima is illustrated in very sharp relief by the appended diagram 2. We draw the following conclusion: the following two vegetative periods were of the greatest importance -- were of critical importance -- for the yield of rye: that of the second half between the beginning of the shooting and the beginning of the heading -- that is, the period immediately before the heading; and the ripening period. But the favorable passage of these phases in the life of the rye plant requires, firstly, good rainfall during the preceding periods, to build up sufficient moisture stocks in the soil by the beginning of the critical phases, and, secondly, the absence of atmospheric aridity, which acts to stunt and depress the growth of the plant in its critical phases. (The direct action of rainfall during the blooming period itself can only be negative, for it exerts a harmful influence on the course of the blooming process. However, under the conditions of the arid climate of the former Samara guberniya, its positive action for the following period -- that of ripening -- overbalanced its immediate harmful effect on the blooming process. When we turn these phases

critical (the second half of the shooting phase and the ripening period) we have in mind the great importance that they had in the life of winter rye from 1885 to 1915, in comparison to the other phases; but we do not by any means mean to imply that these two phases were of decisive importance. Meteorologically there are four phases that are important for these two, namely the fourth period (for its rainfall), the fifth period (for its saturation deficit). The weak influence of the meteorological factors during the period of heading is explained first of all (and primarily) by the fact that during this period the process of the plant-body formation is merely completed, whereas this body is already almost entirely organized by the beginning of heading: not only is the number of spikes determined by the beginning of heading, but also the number of racemes in each spike.

In the second place, the opposing influence of temperature and rainfall are intensified; high temperatures and dry weather act unfavorably on the development of the plant, but, on the other hand, they encourage the smooth heading of the rye and thus prevent the so-called multi-layer condition and safeguard the crop from lodging, or crinkling down (POLEGANIYE) which is already possible even at the heading stage. Finally, we must also remember the importance that the rainfall during the heading period may have had for the following period of blooming, during which the oxygen requirements of the root system are very high, and therefore rainfall during the heading period, which moistens the soil up to the beginning of blooming, thereby impeding the process of soil aeration, creates conditions not entirely favorable for the blooming period. All of these mutually contradictory effects of the weather during

the heading period alternate the effect of the meteorological factors.)

The coefficient of multiple correlation that corresponds to the regression equation in which the yield of rye is regarded as a function of the saturation deficit of the several periods attains the value of -0.918 , as we have seen. When the systematic error is eliminated, this coefficient is reduced to -0.888 .

If we combine the results of the action of the rainfall on the rye crop and that of the saturation deficit into a single joint equation, we obtain a regression equation with a correlation coefficient of 0.972 , which after elimination of the systematic error becomes 0.940 . (The combined action of these weather elements is determined in the following manner.

First we find the general index for each of these factors. This is proportionate for each year to the coefficient in the corresponding regression equation. (The corresponding "weights" for Equation I are given on page 137). We denote the general index by the symbol N for the rainfall and E for the saturation deficit. The general index N is correlated with the yield q (we recall that q represents the index of yield after eliminating the influence of the condition of the sowings on emergence from the snow) to the extent of 0.956 , and the index E is so correlated to the extent of -0.918 . The combined influence of these generalized indices N and E is expressed by the correlation coefficient 0.972 . But this coefficient includes a systematic error, which is eliminated in the following way. The corrected correlation coefficient between N and q is 0.929 , the corrected coefficient between E and q is -0.888 .

the heading period alternate the effect of the meteorological factors.)

The coefficient of multiple correlation that corresponds to the regression equation in which the yield of rye is regarded as a function of the saturation deficit of the several periods attains the value of -0.918 , as we have seen. When the systematic error is eliminated, this coefficient is reduced to -0.888 .

If we combine the results of the action of the rainfall on the rye crop and that of the saturation deficit into a single joint equation, we obtain a regression equation with a correlation coefficient of 0.972 , which after elimination of the systematic error becomes 0.940 . (The combined action of these weather elements is determined in the following manner.

First we find the general index for each of these factors. This is proportionate for each year to the coefficient in the corresponding regression equation. (The corresponding "weights" for Equation I are given on page 137). We denote the general index by the symbol N for the rainfall and E for the saturation deficit. The general index N is correlated with the yield q (we recall that q represents the index of yield after eliminating the influence of the condition of the sowings on emergence from the snow) to the extent of 0.956 , and the index E is so correlated to the extent of -0.918 . The combined influence of these generalized indices N and E is expressed by the correlation coefficient 0.972 . But this coefficient includes a systematic error, which is eliminated in the following way. The corrected correlation coefficient between N and q is 0.929 , the corrected coefficient between E and q is -0.888 .

N and E correlate between themselves to the extent of -0.869 . From these corrected coefficients the coefficient of multiple correlation is readily found, by the formulas for computing correlations, to be 0.940 .

We recall the difference between the original method and the second method for finding these generalized indices. In the first method we take into account the correlation for each period separately between the rainfall and the index of the atmospheric aridity (or temperature), but do not take the mutual correlation between the periods into consideration. By the second method we take into account the correlation between the periods, but do not do so with the correlation between the factors during the separate periods, since this is finally taken into account in its aggregate form, as the correlation between the generalized indices N and E.)

(See diagram 2 on following page)

Let us see how these regularities we have noted are justified in the northern region and in the southern region of the former Samara guberniya. As we have already pointed out, these two regions differ considerably with respect to climate, to the dates on which the several phases of plant development begins, and to the level of productivity. The correlation coefficients between the fluctuations of the rye yield and the rainfall (h) during the several vegetative periods are as follows:

Diagram 2

CORRELATION COEFFICIENTS BETWEEN THE YIELD OF RYE AND THE RAINFALL
(h) AND SATURATION DEFICIT (E - e) DURING THE SEVERAL VEGETATIVE PERIODS

RAINFALL:

Coefficients of variables in regression equation --

Regression Equations --- $q = 5.9766 + 0.2452 p_3 + 0.1506 h_2 + 0.2989 h_2$
 $+ 0.3004 h_4 + 0.02770 h_5 + 0.0106 h_6 + 0.5040 h_7$
 $+ 1.3059 h_8 + 0.2033 h_9; R = 0.929,$

Saturation Deficit:

--- $q = 9.4978 - 0.0475 (E - e)_3 - 0.8493 (E - e)_4$
 $- 0.9003 (E - e)_5 - 0.1156 (E - e)_6 - 0.1171 (E - e)_7$
 $- 0.5473 (E - e)_8 - 0.1192 (E - e)_9; R = 0.918;$
 $R_1; 0.868$

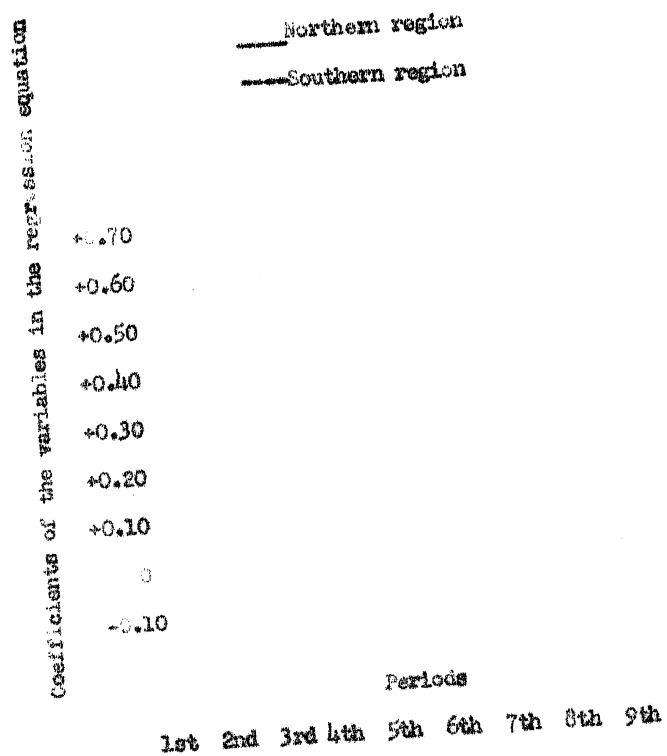
+1.40
+1.30
+1.20
+1.10
+1.00
+0.90
+0.80
+0.70
+0.60
+0.50
+0.40
+0.30
+0.20
+0.10
0
-0.10
-0.20

2nd 3rd 4th 5th 6th 7th 8th 9th
II e p n e bl

178

DIAGRAM 3

COEFFICIENTS OF CORRELATION BETWEEN YIELD OF
RYE AND RAINFALL (h)



CORRELATION COEFFICIENT	NORTHERN REGION	SOUTHERN REGION
$r_{q,h1}$	-0.103	-0.054
$r_{q,h2}$	0.143	-0.075
$r_{q,h3}$	0.418	0.582
$r_{q,h4}$	0.678	0.565
$r_{q,h5}$	0.425	0.531
$r_{q,h6}$	0.036	0.216
$r_{q,h7}$	0.494	0.448
$r_{q,h8}$	0.261	0.271
$r_{q,h9}$	-0.076	0.119

It will be seen that in both northern and southern regions the definite influence of the rainfall begins with the third period, then sharply declines in the earing (6th) period, is reactivated with the onset of the bearing period, only to fall once more to a feeble effect in the 9th, or maturation, period. Thus the wavelike nature of the rainfall influence, established by us for the entire territory of the former Samara guberniya, finds its confirmation also when the data is processed region by region (see diagram 3).

For more detailed analysis we next set up two regression equations, one for the North and another for the South, in which the yield of rye q is correlated with the rainfall during the several periods of the development of rye, beginning with the third, and with the winter rainfall (P_w) (The late autumn rainfall (in decades with temperature of 5 degrees and below) and the early spring rainfall (of the first and second vegetative periods), the independent

influence of which may be considered equal to zero, were combined with the winter rainfall.).

Northern region: $q = -4.6898 + 0.9734h_3 + 1.4137h_4 + 0.5743h_5$
 $- 0.2787h_6 + 0.9322h_7 + 0.7414P_1 + 0.1718P_2$
 $R = 0.738$

Southern region: $q = -5.1912 + 0.7705h_3 + 1.4611h_4$
 $+ 0.014h_5 + 0.945h_6 + 0.4324h_7 + 0.3401h_8$
 $- 0.8211P_1 + 0.4714P_2$ $R = 0.727$

These regression equations, which take account of the mutual correlation between the rainfall of the several periods, show very plainly that there is a "wavelike" of the rainfall throughout the course of the spring-summer vegetative period. Diagram 4 vividly demonstrates this. But at the same time differences are also observed in the significance of the rainfall during the several periods in the northern region and in the southern. These differences entirely agree with the difference in the climatic position of the two regions. The greatest difference is observed with respect to the winter rainfall. In the southern region their influence is fairly well marked. In the northern, more humid region, their significance may as a general rule be considered zero (the significance coefficient of P_1 is most insignificant). It is possible that according to the amount of the autumn rainfall, those in winter exert

either a positive influence on the yield(when autumn has been dry) or a negative influence (when autumn rainfall has been abundant). It is likewise possible that in the North a curvilinear dependence of yield on the winter precipitation is observed (showing the influence of both excessive winter precipitation and of too scanty snowfall to be harmful). All this may be disclosed by a special analysis, but no matter how this question is finally resolved, the difference in the resultant effect of the winter rainfall in the northern and southern regions will still remain undoubted. The contrary effect of the rainfall in the northern region in different years, over the 23-year period, is in the last analysis mutually offset, while in the southern region the favorable influence of rainfall is on balance dominant. A second difference which is likewise explained by the difference between the climatic readings of the two regions, relates to the sixth, or heading, period: the decline in the wave of positive influence of rainfall, which is observed in both North and South, ends, in the South, with a very feeble positive effect during the heading phase, while in the North this fall concludes with a weak negative influence of the rainfall during the heading phase (see footnote to page 140 for explanation). A third difference is the shift, in the southern region, of the maximum influence of rainfall from the fourth period to the third period, nearer to the date when the winter sowings emerge from the snow. Obviously the smaller ground moisture stocks in the southern region are more rapidly exhausted and demand their replenishment from the rainfall. Finally, a fourth difference, clearly demonstrated by Diagram 4, consists in the varying amplitudes of the "wavelike" movements of the rainfall values during the several vege-

tative periods.

In the South the maximum rainfall effect for the several periods is not so sharply expressed as in the North, and this influence is more uniformly distributed over the entire course of the spring-summer life of the rye plant. The following Table 8 will convince us of this. It gives the "weights" for the rainfall values as computed from the coefficients of the corresponding regression equations:

[See Table 8 on following page]

If the influence of rainfall on the fluctuations of the yield of rye were uniform for all 8 periods, then the "weights" for these periods would also be the same for each period "weight" would be equal to plus 0.125 (which would be equivalent to summation of the total rainfall and to the finding of the correlation coefficient between the yield and the sum of these rainfalls), and for this reason the deviations of the "weights" established by us from +0.125 constitute an index of the varied effect of rainfall in the individual periods. These deviations are given in the second part of the table presented above, and the sum of their absolute values for the north and for the south constitutes an index of the degree of non-uniformity in the effect of the rainfall during the several phases of plant development. This index for the north is considerably higher than that for the south: 0.640 against 0.522.

The more uniform influence of rainfall during the course of the spring-summer life of winter rye in the south is explained by the difference between the climatic position of the north and the south:

TABLE 8

<u>Periods</u>	<u>Northern Region</u>		<u>Southern Region</u>		<u>Deviations of the "weights"</u>	
	Coefficients of "WEIGHTS"		Coefficients of "WEIGHTS"		from + 0.125	
	the variables		the variables in		North Region	South Region
	in the regression		the regression equations			
	equation					
Winter Period	+0.0289	+0.018	+0.4714	+0.149	+0.107	+0.024
3rd period	+0.9734	+0.198	+0.7095	+0.224	+0.073	+0.099
4th period	+1.4131	+0.288	+0.4601	+0.146	+0.163	+0.021
5th period	+0.5743	+0.117	+0.3184	+0.101	-0.008	-0.024
6th period	-0.2787	-0.057	+0.0945	+0.030	-0.182	-0.095
7th period	+0.9332	+0.190	+0.4332	+0.137	+0.065	+0.012
8th period	+0.3432	+0.070	+0.3409	+0.108	-0.055	-0.017
9th period	-0.3043	-0.062	-0.3311	-0.105	-0.187	-0.230
Total	(8,9097)	1.000	(3,1591)	1.000	(0.840)	(0.522)

The smaller amount of rainfall and a higher temperature (and likewise the larger saturation deficit) also determine the more rapid consumption of soil moisture stocks in the southern region, and the plants, in consequence, more frequently experience the need for rainfall than in the north. The inferior soil structure in the south (chestnut soils, with a substantial admixture of semi-desert soils in the southeast part of the southern region) accelerate this consumption of soil moisture stocks and thus intensify the action of the meteorological factors.

The common element in the action of rainfall on the fluctuation of the yield of rye in the north and south regions of the former Samara guberniya is expressed not only in the fact that the variation in the effect of rainfall in the individual periods has a wavelike character with two maximums (one during the period of formation of the plant-body, and the other during the period of formation of the grain), but also in the fact that the integral effect of the rainfall during the spring and summer period and the winter period is high both in the north and in the south and is characterized by almost identical correlation coefficients. As we have seen, the coefficient of multiple correlation for the effect of rainfall during the eight periods is 0.938 for the north, which is reduced to 0.907 after elimination of the systematic error, while the corresponding values for the south are 0.927 and 0.890 (for detail see appendix 2).

It is interesting to note that processing of the data for the whole guberniya gave a somewhat higher correlation coefficient;

the corrected correlation coefficient for the guberniya as a whole, as we have seen is equal to 0.929.

It might have been expected that this would be the other way around, since in working up the data by regions the details of the dependence of the crops on the meteorological factors is better taken into account. However, the circumstance that the data for the guberniya as a whole on yields and meteorological characteristics of the weather are more exact or rather "more stable" (editor's note), than the regional data, but in our case the superiority of the regional analysis of the materials outweighs this factor (The possibility of this "overbalancing" is given by the fact that there is much in common between the action of the meteorological factors in the northern region and that in the southern regions which point has been adequately covered in the above discussion. If there were large differences in the degree of the meteorological elements, or, a fortiori, in their character, there could of course be no thought of such "overbalancing".)

It should also be kept in mind that there are some factors which are of secondary importance in comparison to those included in the analysis, who do not enter into our working up, and are not uniformly distributed by territory. The action of agricultural pests, for instance, or parasitic disease, hail, frosts, and to a considerable degree that of dry desert winds as well, belong to this group of factors.

The action of this factor is better aligned by years on large territories than on small ones and for this reason they disturb the established regularity of the basic factors less over extensive

territories than over limited areas.)

With this we conclude our discussion of the papers on the investigations into the dependence of the fluctuations in the yield of rye on the peasant prerevolutionary farm in the arid Povolzh'ye. One of these works is by S. Khorizemenov, and the other is by A. M. Protasov. Both concern the territory of the former Saratov guberniya. A third paper (unpublished) is by me and was done on the former Samara guberniya on a scale incomparably more detailed and accordingly, the conclusions have a far richer content.

We traced down the influence of the meteorological factors on the fluctuations in the yield of rye on the peasant farm of the prerevolutionary epoch, which was distinguished by a very low level of cultural practice. The influence of this level was tremendous. Let us see how the action of the weather was reflected in the fluctuation of yields on the lands of the private landowners. Tracing down this influence, we established what action the meteorological factors had on the fluctuation of the winter rye yield in the arid Povolzh'ye under the conditions of the best cultural practice, since as the farms of the private landowners stood on a higher level than the peasant farms did. As may be seen from the table presented below, the productivity on the land of the private landowners was notably higher than that on the peasant farms while its variability in percent of the mean productivity was somewhat lower.

[See table on following page]

MEAN 23-YEAR PRODUCTIVITY OF RYE IN THE FORMER SAMARA GOVERNMENT		VARIABILITY OF THE YIELD			
		STANDARD DEVIATION		STANDARD DEVIATION IN	
		IN TSENTNER PER HECTARE		PERCENT OF ARITHMETIC	
				MEAN	
PEASANT	ESTATES	PEASANT	ESTATES	PEASANT	ESTATES
5.16 tsentner	6.45 tsentner	2.064	2.300	40	35.6

However, the fluctuations about this higher level of productivity proceeded under the influence of the very same meteorological factors, and this influence was not enfeebled as the following table convincingly shows. This table gives the correlation coefficients between the yields and the indices of rainfall during the spring and summer period (P_g) and with the temperature conditions of the same period (T_g).

Correlation coefficient of yield with
the indices of:

	Rainfall (P_g)	Temperature (T_g)
Peasant Farms	0.839	-0.779
Private landowner estates	0.841	-0.772

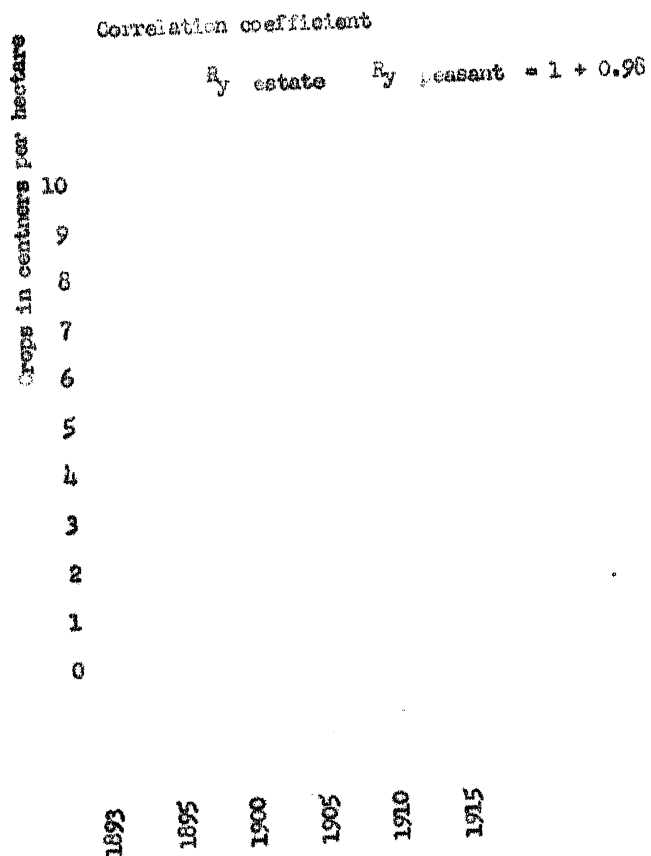
As we see, the correlation coefficients are almost identical.

Diagram 5 vividly demonstrates the very strong parallelism in the fluctuations of the yield on the peasant and landowners lands, which testifies to the fact that the conditions determining these fluctuations were common to both types of farms.

DIAGRAM 5

YIELD OF WINTER RYE ON PEASANT AND LANDOWNER
FARMS IN SAMARA GUBERNIYA

_____ Yield on land of private landowners
 ----- Yield on peasant lands



The correlation coefficient between the yields of peasant farms and private landowner farms is + 0.98.

However the circumstance that these fluctuations occurred at varying levels of productivity, determined by the corresponding levels of cultural practice, led to rather important practical consequences, in spite of the fact that the cultural practices on the farms held by the private landowners was very far from perfection. Table 9 will convince us of this fact.

TABLE 9

Year-groups by amount of spring-summer rainfall P_s Mean yield in year-groups, expressed in tsentner per hectare

	Peasant Farms	Estates
Group I: 5 years with scantiest rainfall. P_s = 1.9 to 7 millimeters	2.08	3.00
Group II: 6 years with rainfall P_s = 9.2 to 12.8 millimeters	5.05	6.19
Group III: 6 years with rainfall P_s = 13.3 to 16.0 millimeters	5.65	7.01
Group IV: 6 years with highest rainfall P_s = 16.9 to 24.8 millimeters	7.27	9.06
23-year mean	5.16	6.45

We see that the data on this table fully confirms the conclusions drawn by us above: 1. The yields both on peasants and on landowners farms varied very strongly in a positive sense, depending on the spring-summer rainfall; (2) but this reaction of yield to the amount of rainfall proceeded at different levels of productivity.

In each group of years with the same rainfall, productivity was therefore higher on the landowners' estates than on the peasant farms. These conclusions agree with all that has been said above. The new element contributed by this table is as follows. Group II of years with reduced rainfall gave a somewhat higher yield on the estates than did group III, with increased rainfall, on the peasant farms; while group III gave a yield on the estates almost as high as that obtained by the peasants in the years with the highest rainfall for the whole 23-year period.

The higher level of agricultural technology practiced on the estates thus led to a notable practical result: these better cultural practices assured the estates of approximately the same yield, under appreciably inferior meteorological conditions, as that obtained by the peasants under definitely better weather conditions, in spite of the fact that the fluctuations of yield on both categories of farms were determined with like force by the meteorological factors.

But the level of agricultural practice on the estates -- which in general was itself far from being perfect -- could still not adequately overcome the influence of a major rainfall deficit (Group I years); and even in years when the rainfall was only re-

duced (Group II years) the yield on the estates did not reach the average level of peasant yield [for the entire 23 years].

The same conclusions follow from Table 10, which is based on the classification of years by their spring-summer temperature indices (T_s). (The difference lies only in the interchange of relative positions between Groups II and III: the estate yield during group II is somewhat lower than the peasant yield during group III, but the estate yield in Group III is higher than the peasant yield in Group IV. The relation is the reverse, as we have seen, for the classification by rainfall.

TABLE 10

YEAR-GROUPS BY T_s	MEAN YIELD BY YEAR -GROUPS (IN TSENTNER PER HECTARE)	
	Peasant farms	Private landowner estates
5-years with highest T_s : from 20.3 to 22.2	2.38	3.30
6 years with T_s from 19.1 to 20.0	4.65	6.11
6 years with T_s from 18.3 to 19.0	6.49	7.86
6 years with lowest T_s : 17.1 to 18.3	6.66	8.01

We continue our comparison by presenting, for the guberniya along the Volga, the mean yields on peasant farms and private estates and the standard deviations in puds per desyatin and in percent of the mean yield. The trend of productivity has been taken into account in calculating the standard deviations (that is, they have been adjusted to allow for the increase in productivity. See Table 11)

Just as in the former Samara guberniya, we have in the estates on a.1 the other Volga guberniyas, in comparison to the peasant farms: (1) a higher productivity; (2) almost the same (but somewhat higher) absolute variability (in puds per desyatin); (3) a lower relative variability of yield.

[See Table 11 on following page]

This difference in relative variability in favor of the estates, due to their better cultural practices, shows up more or less appreciably in two guberniyas, Simbirsk and Saratov, the data of which we now pause to analyze.

We trace the reaction of the winter rye yield to the temperature conditions of the second half of spring and summer (The temperature conditions are treated as follows: (1) the temperature indices are taken from the readings at 1300 hours, since these register the effect of atmospheric aridity better than the noon temperatures; (2) the groupings are made and the correlation coefficients computed from the 1300 hours temperatures for the second and third months after the beginning of the energetic spring growth of winter rye, which

TABLE 11

Former Guberniya	Crop	Mean Yield over 33 years (1883-1915)		Standard Deviations In Puds per Desyatin In Percent of Mean Yield			
		in puds per desyatin		Peasant		Peasant	
		Farms	Estates	Farms	Estates	Farms	Estates
Astrakhan	Rye	13.2	17.0	8.26	9.67	62.7	56.9
	Spring Wheat	20.9	25.8	-	-	-	-
Saratov	Rye	38.0	51.9	12.97	13.89	34.1	26.8
	Wheat	26.7	36.2	11.56	13.80	43.2	38.1
	Oats	32.9	45.7	15.84	18.94	48.0	41.4
Penza	Rye	46.8	57.2	15.97	16.71	34.1	29.2
	Oats	35.6	45.2	15.86	17.12	44.5	37.9
Simbirsk	Rye	45.0	54.3	16.94	16.77	37.6	30.9
	Wheat	33.3	36.8	14.07	14.75	42.2	40.1
	Oats	35.4	42.0	17.20	18.45	48.6	43.9
Kazan'	Rye	42.7	54.5	15.07	18.19	35.3	33.4
	Wheat	32.3	36.6	12.50	13.73	38.7	37.5
	Oats	35.8	42.3	16.12	18.47	45.0	43.7
Ufa	Rye	46.0	55.2	15.95	17.24	34.7	31.2
	Wheat	38.3	42.4	13.98	15.18	36.5	35.8
	Oats	41.1	48.5	19.82	20.91	48.2	43.1

point of time is taken as the first decade with mean diurnal temperature of 5 degrees or higher. For instance, in Saratov Guberniya, in 1893, the onset of spring was in the second April decade (the first April decade with temperature over zero following the last decade with temperature below zero). The first decade of May had a mean diurnal temperature of 5 degrees and above. Thus the second month after the energetic growth of rye began embraces all the June decades, while the third month takes in the July decades. In 1906 the third March decade was the first decade of spring. The second April decade was the first decade with temperature 5 degrees and above. Consequently the second 1906 month after the beginning of spring growth includes the second and third May decades and the first June decade, while the third month includes the second and third June decades and the first July decade.

The temperatures for the former Saratov guberniya were determined from the "Nikolayevskoye" station, located 55 kilometers northwest of Saratov, and those for the former Simbirsk guberniya from the station at Ul'yanovsk (Simbirsk).

We have already pointed out that the year-to-year temperature fluctuations are characterized well enough by the readings of a single station (see Appendix 3).

It is dangerous to use the observations of one station for recording the amount of rainfall, since rainfall is a phenomenon distinguished by very variegated and non-uniform incidence. This circumstance compelled us to confine our analysis of the dependence of winter rye yields on the peasant farms and estates of Saratov and Simbirsk guberniyas to that on temperature fluctuations. To take

account of the influence of the rainfall, a rather large amount of work would be required to compute the rainfall from the observations of several stations. Moreover, as we have already pointed out, the temperature conditions speak not only for themselves, but also to some extent for the rainfall conditions as well, since the two elements are correlated.

We observe that while it is true that our method of tabulating the meteorological factors by movable months is less accurate than that based on the short vegetative periods assigned to the basic phases of plant development, which I used in analyzing the Samara guberniya data, it is still undoubtedly superior in this respect to the usual method of tabulating the meteorological observations by immovable calendar months.) on the peasant farms and landowners estates through the 30 years from 1886 to 1915 inclusive.

[See Table 12 on following page]

The fundamental conclusions established by us on the materials of the former Samara guberniya are confirmed by the data in Table 12 on Simbirsk and Saratov guberniyas. We note first of all that the spring-summer temperature conditions are strongly reflected in the rye crop fluctuations on both peasant farms and landowner estates (for the correlation coefficients for Simbirsk guberniya are -0.73 and -0.76 respectively, while for Saratov guberniya both of them are -0.76). We also see that in Simbirsk guberniya the group of years with increased temperatures (Group II) gave the same yield (even a little larger) on the estates as was obtained by the peasants in Group III years, with average temperatures. This middle group in turn assured the estates of a yield somewhat higher than

TABLE 12

CORRELATION COEFFICIENTS BETWEEN THE YIELD OF WINTER RYE AND
THE MEAN TEMPERATURE AT 1300 HOURS

Former Simbirsk Guberniya		Former Saratov Guberniya	
Peasant farms	-0.73	Peasant farms	-0.76
Estates	-0.76	Estates	-0.76

GROUPS BY MEAN TEMPERATURES AT 1300 HOURS	MEAN YIELD IN PUDS PER DESYATIN	
	<u>Peasant Farms</u>	<u>Estates</u>

Former Simbirsk Guberniya

I. 6 years with highest temperature (22.9-25.4)	27.6	35.0
II. 6 years with temperature 22.0-22.6	40.4	49.6
III. 6 average years with tem- perature 20.8-21.8	47.3	56.4
IV. 6 years with temperature 19.7-20.1	50.9	60.7
V. 6 years with lowest tempera- ture (17.9-19.2)	58.8	68.7

Former Saratov Guberniya

I. 6 years with highest tempera- ture (25.1-26.2)	26.7	39.3
II. 6 years with temperature 23.4-25.1	29.9	45.5
III. 6 average years with tempera- ture 22.7-23.2	39.0	55.7
IV. 6 years with temperature 21.8-22.5	45.0	56.7
V. 6 years with lowest tempera- ture (18.5-21.4)	51.2	67.0

Note: The sign is used to join the year groups with approximately equal peasant and estate yields.

the peasant yields in Group IV years, with more favorable temperature conditions. Finally the estate yields in Group IV years correspond with the peasant yields in Group V years, which had the most favorable weather conditions of all. Here as in Samara guberniya, Group I, with temperature conditions extremely unfavorable for the yield, stands alone: the cultural practices of the estates were unable to overcome these conditions to a sufficient extent and thus raise their yields even to the level of the peasant yields in Group II years.

The influence of cultural practice stands out even more vividly in Saratov guberniya. Here even the Group I years, with the most unfavorable temperature conditions of all, do not stand apart from the other years: for the productivity on the estates, even in this group of years, still equals the peasant productivity in the years of average temperature conditions. The equality of yields between estate and peasant farms "eclipses" a whole group. For instance, the estate yields at average temperature conditions is even a little higher than the peasant yields under the best temperature conditions. Obviously, this effect of cultural practices, (even where the action of the weather on the yield fluctuations about the increased level still remains in full force and effect) should be expected to stand out even more sharply as the level of cultural practice rises.

The farms at the experiment stations did stand on this still higher level of agrotechnology. The materials on rye yield fluctuations at two stations located on the borders of the arid Povolzh'ye were worked up. The first of these projects was carried out

by Professor D. I. Shehukin for the Bezenchuk experiment station, located within the former Samara uyezd ("Atmospheric precipitation as a factor in productivity under the conditions of the steppe Povolzh'ye"). The second work, by P. N. Konstantinov, relates to the Krasnokutsk agricultural station in the south of the former Novouzenek uyezd ("The agricultural value of atmospheric precipitation in the region of the Krasnokutsk agricultural experiment station", Zhurnal opytnoy agronomii Yugo-vostoka (Journal of experimental agronomy in the Southeast), 1927, volume IV, fascicle II). Unfortunately the analysis of the dependence of the rye yield on meteorological factors is very schematic and incomplete in both of these papers. In the first place, only the dependence of yield on rainfall is traced, and the influence of atmospheric dryness (high temperatures and saturation deficits) is left untouched. In the second place, the meteorological factors are allocated to calendar months -- a method, as we have seen, which is rather coarse and insensitive. The length of the observation periods is also small: 20 years for Bezenchuk station (1905-1924), 15 years for Krasnokutsk (1912-1926). In both cases cultural practice was high-level: rye was sown on an early, April fallow. Konstantinov's paper fails to mention the productivity of rye at Krasnokutsk experiment station, but at the Bezenchuk station it attained 96 puds per desyatina (14.39 tsentner per hectare) during the 20 years, which was more than double the average peasant productivity of 40.5 puds per desyatina in the former Samara guberniya, on which Bezenchuk station is located. The productivity of Bezenchuk station is thus 237 percent of this peasant productivity. Professor Shehukin computes the correlation coefficient between the yield of rye at Bezenchuk station

and the May rainfall to be +0.61. We search in vain in the paper for the corresponding coefficients for the other months, but find only the statement that their value for Autumn (August and September) is insignificant. None the less, a combination of the autumn and May rainfall shows that the autumn rainfall also influenced the rye yield at Bezenchuk station, but much less actively than that of May. Specifically Professor Shchukin notes in his paper that in the catastrophic year 1921 the spring-summer drought coincided [i. immediately followed] with the autumn drought, and the rye yield at Bezenchuk station was the lowest of the whole 40-year period, namely 20 puds per desyatin. (3 tsentner per hectare).

Konstantinov's work is more detailed. The highest correlation coefficient between the rye yield at Krasnokutsk experiment station and the monthly rainfall are for April (+0.726) and May (+0.685).

The shift of the maximum correlation coefficient from May at Bezenchuk station to April at Krasnokutsk station (located in the South of the southern region of the former Samara guberniya) entirely harmonizes with the conclusions of my own work on this region: in the southern part of the arid Povolzh'ye the spring vegetation of rye begins earlier, while the more arid climate leads to the more rapid consumption of the soil moisture stocks by the beginning of spring and is consequently also responsible for the earlier need for spring rains. The very appreciable influence of the winter rainfall, likewise in precise agreement with my own work on the peasant farms, is also noted in Konstantinov's paper. He obtained the following results by striking the

general "value-balance of atmospheric precipitation" by seasons, using the algebraic sum of the monthly coefficients of correlation between rainfall and rye yield:

<u>Seasons</u>	<u>Algebraic sum of correlation coefficients</u>	<u>Seasons</u>	<u>Rainfall in Fal- lowing year (12 years of observations)</u>
Autumn (8th-10th months)	0.102	Autumn	0.020
Winter (12th-3rd months)	1.010	Winter	1.532
Spring (4th-5th months)	1.413	Spring	1,210
Summer (6th-7th months)	0.801	Summer	0.013
Total for year		3.326	Total for Year 2,835

As we see from this table, the spring rainfall was of greatest importance on the fields of the Krasnokutsk experiment station, while the winter rainfall stands in the second place and the importance of the autumn rainfall was very small.

It is interesting to note that Konstantinov succeeded in establishing the influence of the winter and spring rainfall of the following years on the rye yield -- that is, during the early spring fallow. This circumstance should not only increase the productivity of rye at Krasnokutsk station but should also make the yield more reliable, since the yield of rye at this station is subjected to the influence of the meteorological factors of two years (mainly that of the spring and winter periods in both cases). The absence of any yield data at all from Konstantinov's paper, whether the mean values for the period studied, or the data for the separate

years, deprived me of the possibility of determining the specific extent of the productivity increase and of the reduction in the variability of that productivity. Things are somewhat better in the case of the Benzenchuk experiment station; Professor Shchukin does give the mean productivity of rye for 20 years, which as we indicated was 96 puds per desyatin, or 14.39 tsentner per hectare, and the minimum yield is given by him as 20 puds per desyatin, or 3 tsentner per hectare. It may be gathered from these fragmentary data that the variability of the rye yield at the Benzenchuk station still remained considerable, in spite of the very marked increase in the productivity by comparison with the pre-revolutionary peasant productivity. However, the reliability of the yield did, never-the-less increase over this period: the minimum yield at the station was 20.8 percent of the mean yield, while on the peasant farms the minimum yield in 1906 dropped to 0.94 tsentner per hectare, which is 15.6 percent of the 23-year mean of 5.16 tsentner. We are able to judge, to a certain limited extent, from this comparison of the deviations of the minimum yields, of the higher stability in the yield on the fields of Benzenchuk station (But in absolute values, the minimum yield at the Benzenchuk station was lower than this average yield by 11.39 tsentner, and on the peasant farms, by 4.22 tsentner [below the average peasant yield]).

In spite of the schematic nature and the incompleteness of the papers of Professor Shchukin, and of Konstantinov (The schematic character and the incompleteness of the Shchukin and Konstantinov papers however, make it impossible to fix the extent of this influence quantitatively. The highest correlation coefficient in the work on the Benzenchuk station is that given between the yield and the May rainfall, and it is +0.64. But this is far from being the

limit of the influence of the meteorological factors. It may very well be that the rainfall for other months give lower correlation coefficients with the yield, but it can be said in advance that if we combine the influence of these factors and of the autumn rainfall and temperature condition with the effect of May the rainfall (using the method of least squares for this purpose), we would obtain a correlation coefficient for the combined influence of all these factors which was considerably higher. The use of the method of short vegetative periods, as has been shown by my experiment in analyzing the data for the former Samara guberniya, is likewise considerably more complete and more accurate in taking into account the influence of meteorological factors. It must be regretted that Professor Shchukin did not give the data by separate years in the paper. In this case a supplementary processing could have filled in the gap in his paper, and, specifically, would have verified whether the same "wavelike" character of the influence of the spring-summer weather, established by me for the pre-revolutionary peasantry with its low level of cultural practice., also held true for the rye yield on the fields of Bezenchuk station.

These gaps in Professor Shchukin's paper are all the more to be regretted, in that the Bezenchuk station obtained incomparably better results with rye, with spring wheat: The mean productivity of rye at the Bezenchuk station is almost $2\frac{1}{2}$ times higher than the mean pre-revolutionary peasant yield, while the same station attained a productivity with respect to spring wheat of only 70 percent higher than that of the peasants, as we shall see.

The highest correlation coefficient for the Krasnokutsk

station is $+0.728$ for the April rainfall, but this is not the limit of the quantitative significance of the meteorological factors. The same remarks which I have made about professor Shchukin's work are also applicable to Konstantinov's paper.

Specifically, the combination of the April rainfall with that of May is something that obviously suggests itself, for both these coefficients of correlation are very close together -- $+0.728$ for April, and $+0.685$ for May.) we see that the improvement in cultural practice at the Besenchuk and Krasnokutsk experiment stations, which was responsible for the very marked increase in productivity, still did not free the fluctuations of yield around this higher level from the action of the meteorological factors, and, above all, from the effect of the spring rainfall, to which must also be added, for the south of the Trans-Volga area, of the winter rainfall as well.

In any case the autumn rainfall is not of primary importance for the yield of rye in either of these papers. In a wake of Protasov, of S. A. Kharisomenov, and myself lead to the same conclusion.

The conclusions drawn by Professor P. I. Brounov in his work "Field crops and the weather" (1912) on the predominant significance of the autumn rainfall in the region of the former Saratov guberniya are thus controverted by all that has been pointed out in the above works (The work of Professor Brounov is based on the material of the Marinsk agricultural farm of the school of agriculture, which is located 55 kilometers northwest of Saratov. Professor Brounov, however, had neither the methods of mathematical statistics nor the elementary statistical classification with these data.

He draws this conclusion exclusively on the basis of these visual impressions of that are given by the diagram he has drawn up. This is entirely inadequate for their categorical affirmation of the preponderant importance of the autumn rainfall even on the field of their Marinsk farm. In this respect, the paper of Professor Brounov unconditionally requires still further supplementary analysis of the material.

APPENDIX I

We present a summary of all statistical data used in this work: (1) data on the rye yield (2) the condition of the rye sowings before their emergence from the snow, and after, (3) the dates of main periods in the vegetation of rye and the time when the sowing of the winter crop began.

1. The data on productivity is taken from the materials of the former Central Statistical Committee. The mean productivity by guberniyas have been determined according to their mean values in the northern region of the former Samara guberniya (Bogursk, Bugul'minsk, Stavropol', Samara, and Buguluk uyezdy) and the southern region, including Mikulayevsk and Novousenskiy uyezdy, while the indices for the northern region were weighted at 0.7 and the indices for the southern region at 0.3, in computing the mean values for the whole guberniya.

These weights were fixed on the basis on the following considerations: the area sown to rye in the southern region, if the opening, middle and closing years of our period are taken (1893, 1904, and 1915), amounted to 29.6 per cent of the entire area under rye in the guberniya, or, in round numbers 30 per cent, so that 70 per cent of the area of rye sowings should be allocated to the northern region.

The considerable differences between the northern and southern regions of Samara guberniya favor this method of determining the norms for the guberniya: the marked difference in productivity, in the length of the vegetative periods (for instance rye ripens about 10 days earlier in the southern region, while the

sowing of rye on the other hand usually commences 10 days earlier),
and in climate.

[See Table I on following page]

The comparison of the mean productivity for the first 11 years,
1893 to 1903, with the mean productivity for the last 12 years, 1905
to 1915, indicates the stationary condition of the productivity of
rye in the former Samara guberniya during the period under study,
for we have almost identical of 5.28 and 5.02 tsentner per hectare.
There is therefore no need of constructing a regression equation,
to separate the trend component in the variation of productivity
from the random annual fluctuations of that productivity. We are
thus enabled to include the productivity data in the investigation
*Heads on right side (4 columns) repeat. Bottom entry in right YEAR
column is: MEAN

TABLE I

YIELD OF RYE IN THE FORMER SAMARA GUBERNIYA
(In tsentner per hectare)

Years	North Region	South Region	Average for Guberniya (0.7, North 0.3 South)	Years	North Region	South Region	Average for Guberniya (0.7, North 0.3 South)
1893	6.27	2.91	5.26	1905	3.27	0.72	2.51
1894	6.60	3.81	5.76	1906	1.03	0.72	0.94
1895	7.36	3.06	6.07	1907	5.58	2.92	4.78
1896	6.31	6.48	6.36	1908	5.97	1.29	4.57
1897	4.93	1.95	4.04	1909	9.28	3.88	7.66
1898	1.90	1.59	1.81	1910	5.65	2.77	4.79
1899	7.81	3.85	6.62	1911	1.72	1.23	1.57
1900	7.60	5.62	7.01	1912	9.40	2.53	7.34
1901	4.33	1.68	3.54	1913	7.44	6.36	7.12
1902	3.88	3.82	3.86	1914	4.53	3.43	4.20
1903	9.33	4.15	7.78	1915	8.25	7.41	8.00
1904	8.67	3.61	7.15				
				Average	5.96	3.30	5.16

without first subjecting it to this preliminary processing.

Table 2

CONDITION OF RYE SOWINGS BEFORE EMERGENCE FROM SNOW
AND AFTER EMERGENCE FROM SNOW.
in balls

Year	Condition in late Autumn	Condition after emergence from snow	Year	Condition in late Autumn	Condition After Emer- gence from snow
1892/93	3.93	4.00	1904/05	2.12	2.00
1893/94	3.59	3.59	1905/06	3.70	3.85
1894/95	2.75	2.75	1906/07	3.07	3.25
1895/96	3.71	3.67	1907/08	2.70	2.50
1896/97	2.68	2.50	1908/09	2.28	2.59
1897/98	2.87	2.50	1909/10	1.75	1.75
1898/99	3.50	3.50	1910/11	2.82	2.75
1899/00	4.25	4.00	1911/12	3.50	3.50
1900/01	3.10	3.00	1912/13	2.61	2.50
1901/02	2.25	2.25	1913/14	2.25	2.25
1902/03	2.50	2.50	1914/15	2.71	3.67
1903/04	2.00	2.00			

The characteristics were established from the agricultural survey of the former department of agriculture and the data ^{of 2} sensitive statistics.

The following numerical expressions were assigned to the characteristics ("average", "satisfactory": 3.0; "entirely satisfactory": 3.25; "above average": 3.50; "considerably above average", "almost good": 3.75; "good" 4.0; "very good": 4.50; "excellent",

living persons
 "superlative", "the eldest/do no remember such a harvest": 5.00; while
 on the side under average: "mediocre": 2.75; "below average": 2.50;
 unsatisfactory": 2.25; "poor": 2.00; "very poor": 1.50; "extremely
 poor": 1.00. We have indicated here almost all the characteristics
 of the conditions of the sowings that are encountered, and those
 that are met only rarely are assimilated to one of these shown above.
 In the agricultural surveys the characteristics are very often
 assigned to uyezdy or regions (north, south, southwest etc). In
 such cases the general characteristic is computed as the mean of
 these partial characteristics for the uyezdy or regions.

3. The information on the vegetative periods and the times
 of sowing of the winter crops have been determined on the basis of
 the data of the agricultural statistics and of the departmental sur-
 veys above mentioned.

[see next two pages for Tables 3a and 3b]

4. The yield with the condition of the sowings after their
 emergence beneath the snow, we determine the existence of a certain
 slight dependence between the rye harvest and the consequences of the
 winter ~~of the winter~~ weather, but mainly with those of the autumn
 weather.

In the following table we present:

[see next page for table]

The regression equation of this table takes into consideration
 the influence on the yield exerted by the condition of the sowings at

3(a) PRINCIPAL DATES FOR FIXING THE VEGETATIVE PERIODS OF WINTER RYE IN THE FORMER SAMARA GUBERNIYA

YEAR	NORTHERN REGION				SOUTHERN REGION			
	Liberation from beneath the snow	Sowing of Spring cereals	Flowering of rye begins	Rye Harvest Begins	Liberation from beneath the snow	Sowing of Spring cereals	Flowering of rye begins	Rye Harvest Begins
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1893	5/IV	17/IV	15/VI	19/VII	22/III	8/IV	9/VI	12/VII
1894	27/IV	2/V	16/VI	30/VII	18/IV	27/IV	15/VI	23/VII
1895	22/IV	28/I V	25/VI	27/VII	5/IV	18/IV	18/VI	15/VII
1896	23/IV	29/IV	22/VI	27/VII	25/I V	30/IV	19/VI	22/VII
1897	12/IV	22/IV	16/VI	18/VII	5/IV	11/IV	12/VI	12/VII
1898	25/IV	1/V	17/VI	17/VII	15/IV	25/IV	13/VI	10/VII
1899	15/IV	23/IV	18/VI	22/VII	5/IV	17/IV	10/VI	11/VII
1900	20/IV	29/IV	21/VI	25/VII	13/IV	28/IV	14/VI	14/VII
1901	5/IV	13/IV	10/VI	10/VII	1/IV	8/IV	3/VI	3/VII
1902	22/IV	1/V	16/VI	17/VII	11/IV	24/IV	11/VI	11/VII
1903	14/IV	18/IV	17/VI	17/VII	4/IV	13/IV	8/VI	8/VII

3(a) cont'd (Page 2)

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1904	20/IV	27/IV	25/I V	31/VII	14/IV	22/IV	17/VI	20/VII
1905	19/IV	24/IV	18/VI	22/VII	17/IV	21/IV	13/VI	13/VII
1906	3/IV	13, IV	2/VI	1/VII	27/III	8/IV	25/V	21/VI
1907	16/I V	25/IV	19/VI	21/VII	5/IV	15/IV	11/ VI	11/VII
1908	26/I V	4/V	23/VI	24/VII	21/IV	28/IV	14/VI	14/VII
1909	18/IV	29/IV	20/VI	24/VII	6/IV	23/IV	14/VI	15/VII
1910	16/IV	22/IV	16/VI	18/VII	5/IV	13/IV	9/VI	11/VII
1911	14/IV	24/IV	17/VI	14/VII	10/IV	17/IV	11/VI	8/VII
1912	15, IV	28/IV	17/VI	19/VII	7/IV	17/I V	7/VI	11/VII
1913	14/IV	21/IV	20/VI	29/VII	10/IV	15/IV	14/VI	20/VII
1914	23/IV	29/IV	23/VI	24/VII	29/III	15/IV	13/VI	13/VII
1915	10/IV	18/IV	11/VI	19/VII	7/IV	12/IV	8/VI	16/VII
Average								
date	Apr 17	Apr 25	Jun 18	July 21	Apr 8	Apr 18	June 11	July 12

(b)

BEGINNING OF SOWING OF WINTER CROPS					
YEAR	NORTHERN REGION	SOUTHERN REGION	YEAR	NORTHERN REGION	SOUTHERN REGION
1892	August 16	August 31	1904	August 19	August 23
1893	August 22	September 2	1905	August 16	August 26
1894	August 17	September 1	1906	August 21	August 27
1895	August 15	August 27	1907	August 21	September 5
1896	August 17	August 29	1908	August 18	August 28
1897	August 20	September 7	1909	August 21	September 13
1898	August 27	September 3	1910	August 21	September 1
1899	August 17	August 27	1911	August 18	August 29
1900	August 15	September 2	1912	August 18	August 26
1901	August 25	September 1	1913	August 26	September 19
1902	August 26	September 4	1914	August 20	August 29
1903	August 28	September 13	Average date	August 20	September 1

FLUCTUATIONS IN THE PRODUCTIVITY OF WINTER RYE IN SAMARA
 GUBERNIYA AFTER ELIMINATING THE INFLUENCE OF EARLY SPRING
 CONDITION OF SOWINGS ON THE YIELD

Year	Condition of winter sowings after emergence from beneath the snow ()	YIELD		$q-y-y_1$	
		y	y_1		
[1]	[2]	[3]	[4]		[5]
1893	4.00	5.26	5.76	-0.50	-0.50
1894	3.59	5.76	5.54		+0.22
1895	2.75	6.07	5.08		+0.99
1896	3.67	6.36	5.58		+0.78
1897	2.50	4.04	4.94		-0.90
1898	2.50	4.81	4.94		-3.13
1899	3.50	6.62	5.49		+1.13
1900	4.00	7.01	5.76		+1.25
1901	3.00	3.54	5.21		-1.67
1902	2.25	3.86	4.80		-0.94
1903	2.50	7.78	4.94		+2.84
1904	2.00	7.15	4.66		+2.49
1905	2.00	2.51	4.66		-2.15
1906	3.86	0.94	5.69		-4.75
1907	3.25	4.78	5.35		-0.57
1908	2.50	4.57	4.94		-0.37
1909	2.59	7.66	4.99		+2.67
1910	1.75	4.79	4.52		+0.27
1911	2.75	1.57	5.08		-3.51

[FLUCTUATIONS IN THE PRODUCTIVITY OF WINTER RYE IN SAMARA, etc.]

[cont'd]				
[1]	[2]	[3]	[4]	[5]
1912	3.50	7.24	5.49	+1.85
1913	2.50	7.12	4.94	+2.18
1914	2.25	4.20	4.80	-0.60
1915	3.67	8.00	5.58	+2.42

$$y_1 = 3.56004 + 0.55074$$

the beginning of spring. The series y , determined by this question, gives the numerical expression of this influence. The series represents the deviation of the yields from the numerical values of y , and thus expresses the fluctuation of productivity after eliminating the influence of the condition of the sowings in early spring on the yield. Obviously, by investigating this series, we establish a more accurate dependence between the fluctuation in yield and the spring and summer meteorological elements.

APPENDIX 2

By operating with the above noted principal moments in the vegetation of rye, we are enabled indirectly, more or less precisely to determine (with an error of a few days at most), a few short vegetative periods, and at that the most important in the life of the plant.

Knowing the beginning of blooming and the beginning of harvesting, we can easily fix the blooming period, which immediately follows the beginning of blooming, the period of maturation that immediately precedes the harvest, and the period of ripening that lies between them, thus dividing the interval between the beginning of blooming and the beginning of the harvest into three parts of equal length. It is true that these periods are not always actually of equal length, but the difference even in this case cannot be significant, firstly by reason of the immediate contact of both initial and final period with the moments of incipient blooming and beginning of harvest, which we have already fixed, and then because each of these periods has a minimum length below which it cannot drop.

It cannot be imagined that the period of blooming or ripening or maturation should last only a few days. Even under conditions that unusually accelerate the process of grain formation -- which conditions are rather frequently observed in the former Samara guberniya (for instance in 1911, 1906, 1901, 1898) -- the period from incipient blooming to commencement of the harvest still lasts 27-30 days, i.e. each period takes on average of 9-10 days. Finally, the interval between the start of blooming and the beginning of the harvest is not so great that large errors could be made with the indirect method of fixing the blooming, ripening and maturation periods (The full period of kernel formation lasted 44 days only in 1894, in the northern region, while in all other years the interval between the beginning of blooming and the harvest held within the range of 39-27 days). There can be no doubt that the extreme days, which really are days of blooming, falls within the blooming periods fixed by us; other extreme days, which are really days of maturation, again fall in our own periods of maturation; and the intermediate days -- the days of ripening -- are likewise caught by our ripening periods. It is similarly easy to determine more or less exactly the earing period, as immediately contiguous to the beginning of blooming. By counting back 9-12 days from the moment the rye begins to bloom, we arrive at the beginning of earing.

The period from the beginning of the energetic vegetation of rye (as marked by the beginning of sowing the early spring cereals) to blooming has a mean duration of 54 days, and fluctuates within the range of 44-60 days for the northern region and 47-62 days for the southern region. This period is too long to be used as a basis for tabulating statistical material. It includes vegetative periods

of different nature, of which the most important are the earing period and the shooting period. To delimit these periods approximately, the method of short decade intervals, already tried out, may be used (See "A" contribution to the question of determining the yield to be expected").

By dividing the whole of this period into 5 intervals of equal length for the year in question, we obtain short periods of 9 to 12 days each, depending on the total duration of the vegetation of rye, from its beginning to the moment of blooming. The interval immediately contiguous to blooming corresponds accurately enough to the period from the beginning of earing to blooming, since, on the one hand, we already know the time of blooming for each year, and, on the other hand, the period from the beginning of earing to blooming lasts from 9 to 12 days in Samara guberniya; we have no data on the shooting of rye in Samara guberniya. But from the information in the then current agricultural statistics of other guberniyas, this vegetative period begins 2½-3½ weeks before earing. Thus if the sixth interval belongs to the earing, then the beginning of shooting falls approximately at the end of the third and beginning of the fourth period.

In general we obtain 9 short periods, which constitute the starting points of our investigation. They are as follows:

- 1st period: early spring growth of rye;
- 2nd period: commencement of energetic vegetation of rye;
- 3rd period: its continuation;
- 4th period: beginning of shooting;
- 5th period: its continuation;

6th period: earing;
 7th period: blooming;
 8th period: ripening;
 9th period: maturation.

By the use of table 3a, which supplies the base points for fixing the vegetative periods, these may readily be determined for each year.

A peculiarity of the printed meteorological data is its tabulation by calendar decades. The vegetative phases are not, of course, taken account of by this system. Therefore it takes the "disseverment" of the decades to adapt the meteorological elements to the vegetative phases. We do this by dividing them into the proportionate number of days allotted to the composite vegetative phases respectively [i.e., those phases extending over more than a single calendar decade]. This technical method of utilizing the meteorological data is best demonstrated by an example.

Let us take the 1894 rainfall. We shall apportion it in detail among the several vegetative phases in the northern and southern regions.

Northern region -- the 1st period, that of the spring germination of winter sowings, occupies the 5 days from 27 April to 1 May inclusive. It thus comprises 4 days of the third April decade and 1 day of the 1st May decade. 0.7 millimeter of rain fell during the last April decade, which gives 0.07 millimeter per day, and 15.7 millimeter of rain fell during the first May decade, which gives the mean value of 1.57 millimeter per day. Consequently we

consider the amount of rainfall for the entire period to be 1.85 millimeters ($4 \times 0.07 + 1.57$).

A total of 45 days elapsed between the beginning of vigorous growth, in 1894, and the beginning of blooming (see table 3). We thus consider each of the following vegetative periods to occupy 9 days.

2nd period (beginning of vigorous growth in rye). Duration 9 days from May 2 to 10. 15.7 millimeters of rain fell during the first May decade. We thus determine the rainfall during the whole of this vegetative period to be 14.13 millimeters (1.57×9).

3rd period (continuation of 2nd), 9 days from May 11-19. Rainfall during 2nd May decade 1.9 millimeter. We take the total [for the period] as 1.71 millimeters (0.19×9).

4th period (beginning of shooting), 9 days from May 20-28, consists of one day of the 2nd May decade and 8 days of the 3rd May decade (which has 11 days, from May 21-31). The mean rainfall per day is 1.09 millimeters. We determine the rainfall for the period to be $0.19 + 8 \times 1.09 = 8.91$ millimeters.

5th period (continuation of the foregoing), 9 days from May 29 to June 6 (inclusive), consisting of 3 days of the last May decade and 6 days of the first June decade, with total rainfall 20.6 millimeters. We consider the total rainfall for the period to be 15.63 millimeters, ($3 \times 1.9 + 6 \times 2.06$).

6th period (earring phase), 9 days from June 7-15 (inclusive): 4 days of first June decade and 5 days of the 2nd (with 10.5 millimeters).

The rainfall determined for the entire period is 24.89 millimeters
 $(4 \times 2.06 + 5 \times 4.05)$.

To determine the rainfall for the last vegetative periods, we first divide the interval between the beginning of blooming and the beginning of the harvest -- which was 14 days in 1894 -- into three equal periods.

7th period (blooming): 14 days from June 16-29, includes 5 days of the 2nd June decade and 9 days of the third June decade, in which rainfall was 28.4 millimeters. We take the amount of rainfall for the entire period as equal to 45.81 millimeters $(5 \times 4.05 + 9 \times 2.84)$.

8th period (ripening): 15 days, from 30 June to 14 July, consists of one day of 3rd June decade, the whole 1st July decade (rainfall 14.5 millimeters) and 4 days of the 2nd July decade (rainfall 15.3 millimeters). We fix the rainfall for the entire period at 23.46 millimeters $(2.84 + 14.50 + 4 \times 1.53)$.

9th period (maturation): 15 days from 15-29 July (6 days of 2nd July decade and 9 days of 3rd (The last July decade has 11 days) decade, with precipitation 17.2 millimeters). We take the rainfall for this period to be 23.22 millimeters $(6 \times 1.53 + 9 \times 1.56)$.

We proceed in exactly the same way for the rainfall in the southern region (Different vegetative periods, of course, are taken for the southern region. See Table 3a, 1894). As a result we obtain, for the year 1894:

VEGETATIVE PERIODS	Northern Region		Southern Region	
	Number of days	Rainfall	Number of days	Rainfall
1st period : early spring				
growth	5	1.85	9	3.54
2nd period : vegetation be-	9	14.13	10	9.56
3rd period : fore shooting	9	1.71	10	5.20
4th period : shooting	9	8.91	9	4.55
5th period	9	15.63	10	10.70
6th period : earing	9	20.49	10	17.26
7th period : blooming	14	45.81	13	30.83
8th period : ripening	15	23.46	12	15.00
9th period : maturation	15	23.22	13	5.78

In view of the fact that the periods are not always of the same length (The length of the periods, beginning with the 2nd, ranges from 9-12 days; 13-day periods are rarely met, and only once, in 1894, for the northern region, the 7th, 8th and 9th periods have 14-15 days) in the different years, we have reduced the rainfall to a uniform interval in all cases, and calculate it on a 10-day basis, then giving the aggregate for the whole guberniya.

Precipitation in millimeters

VEGETATIVE PERIODS	NORTH	SOUTH	FOR WHOLE GUBERNIYA (0.7, North + 0.3, South)
1st period	3.7	3.9	3.8
2nd period	15.7	9.6	13.9
3rd period	1.9	5.2	2.9
4th period	9.9	5.1	8.5
5th period	17.4	10.7	15.4
6th period	31.7	17.3	27.4
7th period	32.7	23.7	30.0
8th period	15.6	12.5	14.7
9th period	15.5	4.5	12.2

Similar computations were made for the rainfall and temperature of the remaining years, and the numerical indices of the degree of "watering" of each vegetative period place, together with the temperature data, in Appendix 3.

It was necessary to have recourse to the indirect method of determining the degree of moistening and the temperature conditions, since there are no data for the rainfall by days in the printed material of the Central Geophysical Observatory. As a result of such computations, we do get a certain amount of "blurring" of the results and errors of closure, deviation of total from 100% in individual years, since the computations are based on the hypothesis of the uniform distribution of rainfall and temperature throughout the "disembled" decade. This hypothesis is justified on balance over an extended period of years, since there is no reason to assume that

one day or another in a decade has any preponderance in rainfall.

Thus in general we have the right to expect a very considerable smoothing out of the errors of closure. In any case these errors of closure, or lack of alignment, even in the individual cases, are far lower than the large non-alignments observed when operating with fixed calendar periods - even short ones - of which we have presented examples above.

We remark that the defects of these non-alignments are considerably attenuated by the circumstance that the vegetative periods have no very distinct limits: there are some sowings that bloom before the average date of blooming in the given year, while there are also fields that bloom later; while even on the same field all the plants do not simultaneously enter this or that phase of ~~growth~~ growth; and besides this the processes themselves develop by smooth gradations, strengthening from zero to the "flush" and then gradually dying down once more towards the end. Thus a certain "blurring" in our determination of the meteorological elements corresponds to a certain indefiniteness existing in reality in the limits of the agricultural periods.

[see next two sheets for table]

APPENDIX 4

The temperature is not distinguished by that wide territorial diversity that is inherent in the rainfall. The observations of the temperature fluctuations at a single point are an index of its variation over a very extensive area with a radius measured in hundreds of

[1]	[2]	[3]	BASIC METEOROLOGICAL ELEMENTS				[8]	[9]	[10]
			[4]	[5]	[6]	[7]			
Year	Yield in TENTHRS per Hectare	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈
1893	5.26	11.0	14.7	5.3	14.2	1.3	7.4	20.3	20.7
1894	5.76	3.8	13.9	2.9	8.5	15.4	27.4	30.0	14.7
1895	6.07	4.3	3.6	3.6	16.6	28.9	8.2	12.4	18.0
1896	6.36	1.1	4.2	25.0	24.5	7.7	41.2	21.1	26.6
1897	4.04	6.4	14.5	9.9	9.9	4.4	2.0	6.1	15.1
1898	1.81	8.2	2.6	5.5	5.3	4.1	9.0	12.0	4.6
1899	6.62	9.4	11.9	9.0	12.5	9.5	1.6	24.3	16.7
1900	7.01	17.6	7.0	6.4	21.4	12.3	9.3	38.1	17.6
1901	3.54	10.9	3.2	3.0	2.7	7.7	9.5	6.3	5.0
1902	3.86	10.7	18.2	8.6	9.5	12.2	16.5	18.4	12.8
1903	7.78	0.3	0.5	16.2	35.1	12.4	12.4	5.6	3.1
1904	7.15	5.5	4.1	4.6	15.1	24.5	24.5	12.1	18.9
1905	2.51	4.0	7.0	6.8	5.0	13.0	14.9	15.3	16.5
1906	8.94	4.4	4.5	1.3	1.6	1.8	2.4	7.0	20.8
1907	4.78	4.4	5.0	12.3	14.9	8.8	11.2	15.1	3.1
1908	4.57	7.6	17.9	18.0	12.5	10.1	6.4	4.5	7.9
1909	4.79	1.3	6.9	14.7	16.2	18.4	10.1	13.8	18.6

100

Basic Meteorological Elements (cont'd)

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
1910	4.79	1.3	6.9	14.7	16.2	18.4	10.1	13.8	18.6
1911	1.57	12.1	3.9	2.5	4.5	8.8	15.8	7.3	3.0
1912	7.34	7.8	10.3	21.7	17.0	12.2	14.5	15.1	20.9
1913	7.12	9.6	6.3	14.4	27.7	24.8	23.4	15.0	26.6
1914	4.20	11.8	6.9	9.0	19.5	21.7	3.8	4.0	4.9
1915	8.00	6.1	21.5	26.1	16.4	7.3	10.7	22.8	8.3

Same as column 1 in first section

Same as column 2 in first section

10.3	10.9	11.9	3.6	3.3	3.0	17.4	17.8
12.2	9.8	8.6	8.7	8.2	7.7	18.3	18.3
12.4	13.4	14.2	5.9	5.6	5.3	17.9	18.0
20.7	22.4	23.5	7.8	7.2	6.7	18.7	18.9
8.5	9.2	9.6	5.9	5.7	5.4	20.0	20.3
5.8	5.4	5.3	7.0	6.4	5.9	21.8	21.8
10.9	11.1	11.5	8.4	8.2	8.0	18.1	18.2
15.1	16.1	17.6	7.5	7.0	6.6	18.5	18.7
4.8	3.8	3.2	7.5	7.0	6.5	20.2	20.4
12.1	10.7	10.0	6.3	5.8	5.4	20.1	20.2
18.4	24.4	28.7	9.3	8.9	8.6	19.5	19.7

Basic Meteorological Elements (cont'd)

[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
13.3	13.3	13.5	10.8	10.3	9.9	17.2	17.1
8.4	7.0	6.1	8.4	7.9	7.5	19.6	19.8
2.5	1.9	1.7	5.5	5.2	4.9	20.9	21.1
11.6	12.8	13.7	7.6	7.1	6.7	19.7	19.9
12.6	13.3	13.3	10.4	10.4	10.4	18.3	18.5
21.3	24.8	27.7	8.8	8.3	7.9	18.3	18.3
13.8	15.0	15.6	10.5	9.9	9.4	19.7	19.1
6.2	5.2	4.6	8.6	8.2	7.8	21.9	22.2
15.8	16.9	17.3	6.1	5.8	5.6	18.8	19.0
20.1	22.5	24.3	8.5	8.3	8.2	17.6	18.3
13.1	15.3	16.8	7.9	7.5	7.2	19.1	19.1
17.3	17.8	17.8	8.1	7.9	7.7	17.9	18.1

BASIC METEOROLOGICAL ELEMENTS (Right hand section)

[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]
22.2	2.1	5.7	7.6	12.0	16.9	19.9	18.5	20.2	22.1	-0.50
12.2	6.2	12.3	17.8	18.2	17.6	17.6	18.5	16.4	21.6	+0.22
12.9	4.4	8.0	12.7	13.6	14.1	21.5	22.4	21.8	20.1	+0.99
22.2	5.0	11.8	12.9	18.3	18.0	17.5	19.8	19.8	19.6	+0.78
10.5	4.6	7.9	14.9	19.5	20.8	20.2	20.2	20.0	21.2	-0.90
16.9	4.3	10.8	19.7	20.2	18.4	18.3	20.8	24.0	25.9	-3.13
9.4	7.4	9.8	12.2	15.1	16.6	19.2	21.2	19.6	24.2	+1.13
20.8	5.1	11.1	15.1	16.9	16.4	19.3	18.1	20.9	21.2	+1.25
8.0	5.1	10.9	13.5	13.3	16.5	18.8	22.7	25.3	25.3	-1.67
14.5	4.1	9.5	12.7	17.3	18.2	21.3	24.7	21.5	21.7	-0.94
3.3	7.5	11.9	13.3	13.3	16.5	19.7	21.9	23.7	24.4	+2.84
10.4	8.6	13.9	17.1	12.3	13.5	16.4	20.8	20.5	19.8	+2.49
16.1	6.0	11.9	14.5	19.1	20.2	20.5	21.5	19.7	19.5	-2.15
15.6	4.0	7.7	15.9	18.7	19.6	21.5	24.3	23.4	20.6	-4.75
18.3	5.4	10.8	12.3	13.0	14.7	18.4	21.7	25.1	26.3	-0.57
8.6	10.4	10.4	11.8	13.3	11.3	20.6	22.1	20.5	20.8	-0.37
7.9	6.5	12.2	15.6	13.3	15.5	21.6	20.1	21.2	22.5	+2.67

Basic Meteorological elements cont'd (Right hand section)

[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]
7.0	7.8	14.4	16.7	15.5	15.8	19.3	20.3	21.1	23.8	+0.27
7.0	6.7	11.3	13.6	15.1	18.4	18.1	22.8	27.7	27.6	-3.51
15.2	4.9	7.8	11.3	15.2	18.6	22.7	24.3	20.9	18.0	+1.85
37.0	7.7	9.7	8.7	12.9	16.5	17.5	18.7	22.3	21.3	+2.18
5.5	6.0	10.7	17.7	16.2	15.7	18.6	21.5	21.6	21.5	-0.60
11.5	7.2	9.4	11.1	14.4	15.7	18.3	18.9	20.1	22.6	+2.42

17.9	2.50	Regression equations:								
18.2	3.00	2 $q = 8.289585 + 0.24422 P_1 - 0.633682 T_1 + 0.097585$ $r = 0.926$ (Weighted proportionately to correlation coefficient).								
18.0	3.00									
19.0	2.50									
20.3	3.00	$q = 10.419066 + 0.189306 P_2 - 0.703394 T_2 + 0.08503$ $r = 0.935$ (Weighted proportionately to square of correlation coefficient).								
21.9	3.00									
18.2	2.50	$q = 10.929685 + 0.160915 P_3 - 0.716399 T_3 + 0.094510$ $r = 0.933$ (Weighted proportionately to cube of correlation coefficient)								
18.7	2.00									
20.4	3.00	$q = 7.603127 + 0.188007 P_2 - 0.650248 T_2 + 0.085375$ $r = 0.743750$ $R = 942$.								
20.2	2.5-									
19.8	2.75	EXPLANATION OF SYMBOLS USED IN TABLE								
		(1) $P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10}$: Rainfall by periods;								

Basic Meteorological Elements, cont'd (right hand section)

[11]	[12]	etc.	Explanation of symbols used in Table, cont'd
17.0	2.75		(2) $t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9$: mean diurnal temperature by periods;
19.8	3.00		(3) q : index of yield, expressed as deviations, after eliminating influence of condition of sowings at emergence from beneath the snow;
21.1	2.00		(4) P_1, P_2, P_3 : generalized rainfall, computed proportional to coefficient of correlation, its square and its cube respectively.
20.0	2.50		(5) $[mean] \bar{T}_1, \bar{T}_2, \bar{T}_3$: the generalized temperature of first two periods, similarly computed.
18.5	2.50		(6) T_1, T_2, T_3 : the same for the temperatures of the remaining periods.
18.3	3.00		(7) M : index of haze.
19.1	2.50		
22.4	2.00		
19.0	2.75		
18.4	2.75		
19.0	2.75		
18.2	3.00		

versts (we have in mind the mean ten-day indices of the mean ten-day indices of the mean diurnal temperatures).

There is thus no need for the largest possible number of observation points for a characterization of temperature conditions. We did not find a single station within the former Samara guberniya, listed in the printed materials of the Central Geophysical Observatory, that had uninterrupted observations covering the spring and summer months throughout the entire 23-year period under study. We, therefore, selected, to characterize the temperature conditions in the northern region of the former Samara guberniya (Bugul'minsk, Buguruslan, Stavropol', Samara and Buzuluk uyezdy), four of the nearest stations: Saratov, southwest of this region, Kazan', to the northwest of it, Ufa, to the northeast, and Orenburg, to the southeast. Each of these stations, by virtue of its position, characterizes the northern region of Samara guberniya on one side, but the ten-day mean temperature indices derived from the observations of all four stations undoubtedly defines the average temperature conditions of this region very well, since it is located within the quadrilateral Saratov-Kazan'-Ufa-Orenburg.

We are completely convinced of this when we compare the mean diurnal temperatures at these four stations with the temperature indices at Samara, for which there are incomplete observations from 1893 to 1915. Samara is west of the northern region and is also fairly far way from Saratov, Kazan', Ufa and Orenburg. The correlation coefficients between the Samara temperature and the mean temperature at these four stations is very high and often indicates an almost functional parallelism.

CORRELATION COEFFICIENTS BETWEEN TEMPERATURE AT SAMARA AND AT FOUR OUTER STATIONS MENTIONED			NUMBER OF YEARS' OBSERVATIONS
DECADES			
April 1st decade	+0.963		15
2nd decade	+0.991		13
3rd decade	+0.988		14
May 1st decade	+0.930		15
2nd decade	+0.980		15
3rd decade	+0.976		15
June 1st decade	+0.973		15
2nd decade	+0.981		17
3rd decade	+0.974		15
July 1st decade	+0.984		14
2nd decade	+0.945		16
3rd decade	+0.976		

The observations at Saratov, which is located on the boundary of the former Samara guberniya, were taken to characterize the temperature conditions of the southern region (the former Novouzensk and the former Nikolayevsk uyezdy). Its position with respect to the north and south of this region is satisfactory. Its western location gives Saratov a certain unilateral quality. It was not possible to avoid this unilateral quality, since in view of the gaps in the observations at the Ural'sk station, which is located somewhat to the east of the southern boundary of Samara guberniya, the attempt to use them as a counterpoise to Saratov had to be abandoned.

APPENDIX 5

The first hypothesis that suggests itself is that the importance of the meteorological elements is proportional to their respective correlation coefficients.

COEFFICIENTS OF PARTIAL CORRELATION BETWEEN		
VEGETATIVE PERIODS	YIELD AND RAINFALL	"WEIGHTS"
In 1st period	0.083	0.0356
In 2nd period	0.268	0.1156
In 3rd period	0.484	0.2091
In 4th period	0.762	0.3291
In 5th period	0.312	0.1346
In earing period	0.213	0.1050
In blooming period	0.163	0.0704
TOTAL	2.315	1.0000

Consequently the mean "weighted" rainfall for the entire period of the spring and summer life of winter rye is defined by the

formula:

$$T_f = 0.0356p_1 + 0.1156p_2 + 0.2091p_3 + 0.3291p_4 + 0.1346p_5 + 0.1050p_6 + 0.0704p_7 \text{ where } p_1, \dots, p_7 \text{ are the precipitation by periods}$$

The mean weighted temperature separately for the periods with negative and positive influence on the yield is determined in exactly the same way by the formula:

$$T_1 = 0.5117t_1 \pm 0.4083t_2$$

$$T_1 = 0.0544t_3 + 0.1892t_4 + 0.2081t_5 + 0.1504t_7 + 0.2125t_8 + 0.1854t_9 \text{ where } t_1 \text{ and } t_2 \text{ are the temperature by periods}$$

The regression equation for the dependence of the rye yield on the "generalized" rainfall and temperatures, defined by the method of least squares, is of the following form:

$$y_1 = 2.2470 + 0.2441P_1 - 0.6337T_1 + 0.0776Z_1$$

The coefficient of multiple correlation that characterizes the degree of influence exerted by the rainfall and temperature on the yield, is $R_1 = 0.926$.

We use the squares of the partial coefficients as a basis for the "weighting". This is equivalent to assigning a relatively greater importance to the rainfall and temperature in periods with high correlation coefficients, and a lesser importance to the meteorological elements of periods with low correlation coefficients. The formula for computing the "generalized" rainfall and temperature have the following form, for the second hypothesis:

$$\begin{aligned} P_2 &= 0.0044P_1 + 0.0667P_2 + 0.176P_3 + 0.5393P_4 + 0.0704P_5 \\ &\quad + 0.0541P_6 + 0.0247P_7 \\ T_2 &= 0.0160E_1 + 0.1144E_2 + 0.205E_3 + 0.1222E_4 + 0.2452E_5 + 0.1905E_6 \\ Z_2 &= 0.1675E_1 + 0.3225E_2 \end{aligned}$$

The corresponding regression equation is as follows:

$$y_2 = 10.4191 + 0.893P_2 - 0.7034T_2 + 0.0850Z_2; R_2 = 0.935$$

Finally, we set up a third hypothesis: the importance of the meteorological factors of each period is proportional to the cube of the partial correlation coefficients, which is equivalent to a still further increase in the "weights" for the periods with increased correlation coefficients.

In consequence we obtain the regression equation:

$$q_2 = 10.1217 + 0.1001 P_1 - 0.1104 T_2 + 0.0945 T_3$$

$$R_3 = 0.933$$

There would be no point to continuing to set up variants based on still other hypotheses, since by increasing the "order" of the coefficients of correlation by weighing the elements, we would obtain smaller and smaller R (R_3 already is R_2), and these hypotheses are equivalent to almost total exclusion of the periods with low correlation coefficients: the hypothesis of "third order coefficients" already assigns a "weight" of 0.708 in the formula to the rainfall (P_3) of the fourth period, and a "weight" of 0.182 to that of the third period, reducing the value of the rainfall in all the remaining periods to the total of 0.110 for all five "weights".

Thus we have:

By the first hypothesis, $R_1 = 0.926$

By the second hypothesis, $R_2 = 0.935$

By the third hypothesis, $R_3 = 0.933$

The most probable system of weighing the rainfall in this case is that which is proportional to the squares of the correlation coefficients, since of all three hypotheses it gives the lowest sum of the squares of the deviations of the observed values q from the values q_2 calculated from the regression equation -- which is indicated by the fact that the value of R_2 is the highest (As is clear from the above, the proposed method of approximate determination of the "weights" of the meteorological elements of the short periods is not arbitrary, but is based in the last analysis on the method of least squares. The distinction is only in that we operate

under our method with systems of weights, selecting therefrom the one that satisfies the requirement that the squares of the deviations from the regression line shall be a minimum. It is not superfluous to add that if we were to set up a regression equation, *large* *artia*, in 18 variables, by the method of least squares, then in any case R could not be less than the greatest of our own three R 's, and it would obviously be inevitable for it to be larger than all the remaining R 's, i.e. $R \geq R_2 > R_3 > R_1$. Thus by selecting the weighting system that gives the highest value of R , we give the regression equation that is closest to the "classical" regression equation (in 18 variables).

It is this variant which we have adopted and used in Table 6 of the text.

APPENDIX 6

Besides these considerations to explain the almost total "annulment" of the positive influence of the autumn temperature on the end results of the yield, the following supplementary hypotheses may be advanced.

The positive influence of rainfall acts on the condition of the winter sowings with especially great force during the period of the sowing and germination, up to the inception of tillering, while the positive influence of temperature during this period is very weak, since there is no lack of warmth at this time. The period of sowing and germination cannot, however, do other than leave its impression on the yield of rye, since the conditions under which the grain germinates and the first growth of the sprouts takes place

affect the degree of crop density in the field, which is a very important element in the future yield. If the germination is poor, and there is partial mortality of the emergent seedlings, the crop density of the field may become thinned out to an extent not susceptible of subsequent correction, i.e. so that the field cannot become less sparsely occupied (if we leave out of the picture the possibility of resowing of the field, or of supplementary sowing of the winter crop). The favorable influence of temperature on the condition of the winter sowings occurs mainly during the tillering phase. Unfavorable temperatures (insufficient warmth) create conditions that retard the growth and tillering of winter sowings, which is reflected in the external appearance of the sowings (stunted growth and inadequate tillering). But this defect (stunting and inadequate tillering) can still be made good with the onset of warm spring weather.

The weakness of the influence of warmth during the early spring period on the end results is due in essence to the self-same contradictory components in the action of temperature during this period; warm weather has a good effect on the growth of winter sowings that have emerged from the snow, yet that same warm weather favors reduction in the soil moisture stocks; cold weather maintains the level of the soil moisture stocks, but does not encourage the spring development of the winter crops. All in all, the balance tips slightly towards the side of the positive influence of warmth over the 23-year period.

The influence of haze likewise proves insignificant throughout the 23-year period. To some extent this is explained by the fact

that it was far from being a regular annual visitation. Throughout the 23 years, serious damage was done by it three times, while appreciable damage was observed seven times. In part, too, this is explained by the fact that the influence of haze is included to a considerable extent in that of high temperature. "Dry" mist, "haze", "ponok" leads to higher atmospheric temperature, since the air is filled with minute suspended particles of solid matter, which heat up strongly in the rays of the sun. The supplementary harmful influence of haze consists in these overheated solid particles coming in contact with the leaves (or if the wind is strong, "bombarding" them) and causing "scorching" in consequence. This is the supplementary hypothesis which explains, inter alia, why damage to the plants is observed, after the haze, in the form of yellow, often reddish specks, of which a whole network covers the leaves.

UNCLASSIFIED
CONTRACT

Chapter 2.

SPRING WHEAT

We shall begin our exposition with the results of my unpublished works on the peasant farms of the former Samara guberniya. These works investigate the question in two variants. In the first variant the fluctuations in the spring wheat yield on the territory of the entire former Samara guberniya are compared with the meteorological factors by decades, or ten-day periods. My investigation showed that as a whole, on the whole territory of the former Samara guberniya, the spring rains, from the third April decade to the first July decade inclusive, exert a tremendous influence on the yield. Wheat sowing usually begins in the former Samara guberniya during the first days of the third April decade, and the harvest commences at the beginning of the first August decade. The rainfall thus showed an influence on the yield of spring wheat from the beginning of sowing to the beginning of ripening, or, in extreme cases, even until its second half. The analysis further showed that the number of days with rainfall of 1 millimeter and over is likewise reflected, to no lesser degree, in the wheat yield. This shows that not only the amount of rainfall but also its more uniform distribution is of importance. This gave rise to the thought of combining these two indices into one. It can be done by using the methods of mathematical statistics, first expressing the rainfall series and the rainy day series in what are termed standard deviations, i.e., in mean-square deviations or "sigmas." A single index is determined for each year as the mean of the indices of rainfall and number of rainy days thus transformed. This combined index, to which the term "rainfall index" may be applied (denoted by the symbol h) gives the following correlation coefficients with the fluctuations of the wheat yield:

PERIODS	Coefficients with		
	Correlation with Rainfall (h)		Saturation Deficit (E-e)
1. 1st April decade	+0.205		no data
2. 1st May decade	+0.357	+0.357	-0.291
3. 2nd May decade	+0.181		-0.021
4. 3rd May decade	+0.598	+0.906	-0.698
5. 1st June decade	+0.394		-0.357
6. 2nd June decade	+0.459	+0.767	-0.247 -0.876
7. 3rd June decade	+0.360		-0.570
8. 1st July decade	+0.319		-0.571
9. 2nd July decade	-0.104		-0.198
10. 3rd July decade	-0.184		-0.170

This table shows that the rainfall has a positive effect for the spring wheat yield from the third April decade (beginning of sowing) to the first July decade inclusive. The total rainfall for these eight decades gives a correlation coefficient of +0.906 with the yield fluctuations, which indicates the strong influence of moisture on the fluctuations in spring wheat yields in the arid Povolzh'ye. This influence begins with the time sowing commences and breaks off with the beginning of ripening or with its second half. Its importance is especially intensified from the third May decade, which corresponds approximately to the beginning of shooting. From this moment the action of atmospheric aridity (the second column in the above table) on the spring wheat yield begins. This action is expressed by a correlation coefficient of -0.876.

The regression equation for the combined influence of both factors (the rainfall for eight decades and the saturation deficit

for five decades) is as follows:

$$y = 75.2548 + 66.1056_1h_8 - 6.2386_4(E - e)_8; R = 0.929$$

where y denotes the actual yield in percent of the "normal" yield for the corresponding year; $_1h_8$ is the total rainfall for the eight decades; $_4(E - e)_8$ the accumulated saturation deficit, beginning with the fourth decade and ending with the eighth. The coefficient of multiple correlation of 0.929 evidences the very strong influence of the spring-summer weather in the fluctuation in the spring wheat yield.

It is entirely appropriate to divide the total rainfall into two periods: the first covering three decades, the third April decade and two May decades; and the second covering the five decades from the third in May to the first in July inclusive. This period corresponds to the period when the importance of the saturation deficit for the spring wheat yield is at a maximum. Thus the second period, which is of the most importance in the life of spring wheat, is simultaneously characterized by two indices of weather conditions: the degree of soil moisture and the degree of saturation deficit of the air. The regression equation for the combined influence of these three factors is as follows:

$$y = 60.5517 + 23.3109_1h_3 + 45.6161_4h_8 - 5.6596_4(E - e)_8; R = 0.93.$$

This increase in the amount of detail in the investigation did not, however, lead to increasing the correlation coefficient R indicating the degree of dependence between yield fluctuations and the combined effect of all factors included in the study. (The inclusion of second-order factors in the study -- the temperature conditions of the first period and the rainfall during the second and third decades of July (the ripening period, or its second half, together with the maturation period) likewise results in no appre-

able increase in the coefficient of multiple correlation R , though it does enable us to note that the temperature conditions during the first period have a weak positive effect, while the rainfall of the ninth and tenth decades shows a weak negative influence: $y = 45.1587 + 24.5519_1 h_3 + 51.5582_4 h_8 - 5.2418_4 (E - e)_8 + 0.6163_1 t_3 - 7.8023_9 h_{10}$; $R = 0.941$.

Some increase in the correlation coefficient can apparently be attained by passing over to shorter periods, for the influence of the rainfall (and likewise, in all probability, that of the saturation deficit) does not remain unchanged throughout extended intervals. The following regression equation, in which the rainfall is given in twenty-day intervals, will convince us of this:

$$y = 59.861h + 20.965h_1 h_2 + 27.1155_3 h_4 + 23.3007_7 h_8 - 9.104h_9 h_{10}; R = 0.929.$$

As we can see, the influence of the rainfall gradually increases to the fifth and sixth decades (the first and second in June), then declines, becoming negative in the last two July decades.

The coefficient of multiple correlation is 0.929.)

Analysis of the dependence of the spring wheat yield in the southern, more arid part of the former Samara guberniya, consisting of the former Nikolayevsk and the former Novouzensk uyezdy, represents the second variant.

As in the first variant, this part of the investigation too is based on the crop data for the thirty-year period from 1886 - 1915 inclusive, i.e., for almost the entire period covered by accurate and monotypic productivity statistics. Only 1883, 1884 and 1885 were left out, owing to the insufficient meteorological

data for these years. Just as with the first variant, the trend component due to the improvement in cultural practice was segregated by the method of least squares from the other variations in the dynamics of productivity, and the fluctuations of the yield above and below this changing trend level were studied -- fluctuations determined almost exclusively by factors of meteorological order and represented by the deviations in the yield from the corresponding "normal" productivity of each year. The first and second variants were both based on the meteorological observations of stations with uninterrupted operation, since the year-to-year dynamics of the meteorological factors is reproduced far better in this case. Moreover, a special analysis showed it to be not only possible, in the interest of increasing the number of stations, but even advisable, to include data from meteorological points located outside the territory under study, but near it. The observations of such stations, however, should be taken with lower weights.

The adaptation of the tabulation of the data of the meteorological stations to the growing season of each year represents a substantial distinction from the first variant. We did not succeed in doing this in detail, as was done with rye, since sufficiently full data on the beginning of blooming or heading of spring wheat (It may be relevant to remark that these two moments are close together with spring wheat.) was lacking, but from the data of the sensitive statistics and the agricultural statistics of the former Department of Agriculture ("... year in Agriculture") it was possible to fix the decades in each year when the sowing and harvesting of the spring cereal crops began. The whole life-period of spring wheat was then divided for each year into nine short vege-

tative periods, among which I then allocated the meteorological data. Those interested in the details of this tabulation are referred to Chapter II, Appendix 1.

The results of the influence of meteorological factors on the fluctuations in the yield of spring wheat in the southern part of the former Samara guberniya, allocated to the short vegetative periods, are given in Table 2.

[see Table 2 on following page]

The following is indicated by these correlation coefficients:

(1) The reaction of the yield to the number of days with rainfall of 1 millimeter and above is positive, as was to be expected. (2) Conversely, the correlation of the yield with insufficient saturation of the air with water vapor is negative; except for the first two periods, with very small positive coefficients, the minus sign for the remaining periods marks the harmful influence of dry air, (and also of high temperatures, which are reflected in the indices $(E - e)$), on the yield of spring wheat in the Trans-Volga. (3) It is a striking fact that the effect of all factors increases towards the middle of the growing season of spring wheat, while for the saturation deficit the sharp rise in the correlation coefficient lags one decade behind, but then, for that, ends later. (4) It is very important to note that there is not even one short period in which the action of any single factor on the spring wheat yield during the period under study can be held to be decisive or even definitely dominant; the highest correlation coefficients are in the range of 0.55 - 0.60. It is obvious that these short intervals each superimpose their own impression on the yield, which is then determined by the cumulative effect of all of them. As appears from Table 2, the number of rainy days from the third to the sixth

TABLE 2

COEFFICIENTS OF CORRELATION WITH THE YIELD

Periods	Number of Days	Rainfall	Saturation
	With Rain (d)	(p)	Deficit (E - e)
1st period: the sowing decade	+0.176	+0.136	+0.286
2nd period: second decade after beginning of sowing (germination)	+0.179	+0.272	+0.116
3rd period tillering, beginning	+0.461	+0.471	-0.107
4th period of shooting	+0.402	+0.398	-0.405
5th period: before heading	+0.597	+0.599	-0.418
6th period decades of heading	+0.552	+0.425	-0.598
7th period and blooming	+0.249	+0.106	-0.573
8th period: second decade before the harvest (ripening)	+0.147	+0.134	-0.389
9th period: first decade before the harvest (maturation)	+0.105	+0.073	-0.241

periods inclusive exerts a special effect on the spring wheat yield; and these periods have the highest correlation coefficients (+0.402 to +0.597). If we add up the number of days from the third to sixth periods, inclusive, which corresponds approximately to the phase from a moment somewhat earlier than the beginning of tillering down to the beginning of blooming, and then determine the correlation coefficient between the number of days with rain and the yield, we find that it suddenly attains the very high value of +0.833. By including the number of rainy days first of the second period, and then also of the first, we increase the correlation coefficient to +0.847 and then to +0.883. Finally, when we include, as well, the number of rainy days in the seventh period, we reach a value of +0.915. The further inclusion of the eighth and ninth periods (ripening and maturation) lowers the coefficient again to +0.885 and +0.830. Thus the number of rainy days from the beginning of sowing to the inception of ripening undoubtedly appears to be one of the decisive factors in the fluctuation of the spring wheat yield in the region under study.

The correlation coefficient between the number of rainy days (with 1 millimeter and over of rainfall) during this period is +0.915. The rainfall for the same combined period likewise gives a maximum value for the coefficient of correlation with the yield. It is +0.841, but, as we see, it is appreciably lower than the corresponding coefficient for number of rainy days. This leads to the thought that under the conditions of the Trans-Volga the frequent wetting of the soil during the period under study is more important than the amount of the rainfall. It must, however, be kept in mind that the number of rainy days, determined from the observations of only a small number of stations, may better represent the

degree of humidity or aridity of the weather than the amount of rainfall determined by the same stations, since the amount of rainfall is distinguished by a wider range of variation over territory than is the number of rainy days; for a random shower at any of the small number of stations chosen for processing may exert an appreciable influence on the mean value of these stations' observations. (It may be that this is likewise to be explained by a curvilinear dependence of yield on rainfall in the region under study. We were not, however, able fully to establish this peculiarity, and I therefore refrain from presenting the results of this part of the work. The dependence of the yield on the number of rainy days is rectilinear.) We shall henceforth operate with the number of rainy days, as a more effective factor. The question arises whether the correlation coefficient would not be increased (and it could not be reduced) by breaking up the period from the beginning of sowing to the end of ripening into two components. Thus, for instance, the segregation period of the first two decades after sowing begins, during which the correlation coefficients are insubstantial, naturally suggests itself.

However, on setting up a regression equation in which one of the variables is the number of rainy days from the third decade after sowing to ripening, we do not increase the coefficient of correlation of the joint influence of these two factors on the yield. In exactly the same way, the variant in which the number of rainy days in the first three decades is segregated as one variable, and the number of rainy days from the fourth decade after sowing to the ripening is the second variable, gives identical results. For both variants the correlation coefficient is 0.918.

There is thus no ground for multiplying unnecessary variables, and we may remain, for the rest of our analysis, with the

number of rainy days from the beginning of sowing to the ripening, without dividing this period into components.

We pass now to the saturation deficit. Proceeding in the same way as with the number of rainy days, we find that the most effective quantity is the sum of the mean saturation deficits at 1300 hours from the fourth decade after sowing down to and including the ripening phase: for the correlation coefficient for this period is 0.814. In contrast to the number of rainy days, this period does not include the first three decades after sowing, but does include the ripening period. Without the ripening period, the correlation coefficient falls to 0.789.

The harmful influence of dry air thus begins to be observed a month after the sowing begins (The accumulated saturation deficit for the first three decades after sowing starts gives a correlation coefficient very close to zero, namely: 0.069.), and continues down to the inception of maturation in spring wheat, thus taking in the period of ripening. This factor is less powerful than the number of rainy days. The correlation coefficient is 0.814 as against 0.915. (Possibly because the period of action by the saturation deficit is shorter.) But its action is highly significant, and we therefore include it in the further analysis. We shall consequently dwell now on only two indices. We set up the regression equation which defines the yield as a function of the two variables ${}_1D_7$, the number of rainy days from sowing to blooming inclusive, and ${}_4(E - e)_3$, the accumulated saturation deficit from the fourth decade after sowing down to ripening inclusive. This equation is of the following form: y (crop fluctuations) = $56.6645 + 8.3047{}_1D_7 - 1.0711{}_4(E - e)_3$; $R = 0.932$

If the number of rainy days is divided between two periods: (1) during the first three decades after sowing, and (2) from the fourth decade to blooming time inclusive, and the regression equation in three variables is then found, the coefficient of multiple correlation will remain unchanged:

$$y = 55.5668 + 7.9944_1 D_3 + 8.4622_4 D_7 - 1.0564_4 (E - e)_8; R = 0.932.$$

This equation has been put into a form corresponding to that of the resultant equation of the first variant, the object of which investigation was the study of the yield fluctuations over the entire territory of the former Samara guberniya (cf. equation II on page 171). As we see, the results obtained are very similar: in both cases the rainfall has a significant effect throughout the entire life of spring wheat with the exception of the last two decades, and in both variants the negative influence of the saturation deficit comes into play in about a month after sowing. The coefficients of multiple correlation are likewise very high and are almost identical. But there is also a slight difference: the influence of atmospheric drought in the first variant breaks off before ripening begins, while in the second variant it continues during ripening as well. We give the preference to the second variant, since it takes account far better of the vegetative periods in the life of the plant.

The basic equation henceforth to be employed by us is thus an equation in two variables:

$$y = 56.6645 + 8.3047_1 D_7 - 1.0711_4 (E - e)_8; R = 0.932.$$

The correlation coefficient is not very sharply increased here by comparison with that for $_1 D_7$ alone.

This is explained by the intercorrelation between D and $(E - e)$: marked raininess is usually accompanied by a lower saturation de-

ficit, and conversely: greater dryness of the air (and higher temperatures) are also accompanied by a smaller number of rainy days. Thus each of these two factors speaks not only for itself alone, but also to some extent for the meteorological factor associated with it. The correlation coefficient of 0.932, considered by itself, is very high and testifies to the fact that these two indices selected by us are really dominating and determine the fluctuations in the yield of spring wheat in the southern part of the Trans-Volga from 1886 to 1915. The question arises whether this coefficient could not be further increased by including new factors. Thus we have not indicated the rainfall of the last two decades: the periods of ripening and maturation. In the very same way the period of the first three decades from the beginning of sowing has also been ignored from the standpoint of temperature. We add these two factors to the regression equation, and obtain the following equation as a result: $y = 68.1287 + 8.1591_1 D_7 - 1.1914_4 (E - e)_8 + 0.3201_1 t_3 - 1.6679_8 D_9$; $R = 0.935$.

As we see from this equation, the temperature conditions during the first month after the beginning of sowing exert a very weak positive influence on the yield of spring wheat (and this is entirely natural), while the raininess of the ripening and maturation periods displays a similarly weak negative action. This agrees with the conclusions that result from the first variant of our work. But the effectiveness of these factors is so insignificant that the correlation coefficient that takes them into account remains almost unchanged (0.935 instead of 0.932).

We carry the investigation further into detail along the line of searching for the curvilinear dependences of the yield on these two basic factors. As remarked above, the yield fluctua-

tions are there in a rectilinear connection between the yield fluctuations and the number of rainy days. As for the saturation deficit, measured by the value $(E - e)$, this was found to exercise a peculiar influence on the fluctuations of the yield of spring wheat in the region under study. This influence was exerted by the saturation deficit from the fourth decade after the beginning of sowing to and including the seventh period, that of blooming. The extreme deviations of the saturation deficit, both negative and positive, show the stronger effect, which is represented by the equation of a third-order parabola:

$$y = 0.8486 - 0.150422 \Delta H_7 + 0.005901 \Delta H_7^2 - 0.004389 \Delta H_7^3; R = 0.841$$

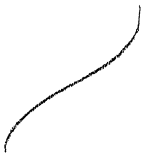


Figure 1

In this equation the symbol ΔH_7 designates the deviations from the thirty-year mean of the saturation deficit during the period from the fourth decade after the sowing to and including the blooming period. This dependence has the following form graphically, representing the increase in productivity as dependent on the reduction of the saturation deficit (Figure 1).

This form of dependence thus evidences the fact that in its extreme manifestations, both positive and negative, the degree of dryness of the air during the period in question acts, as it were, to restrain the influence of the other factors, which enjoy a greater freedom of action when the saturation deficit is more nearly normal. The coefficient of curvilinear dependence is higher than that of rectilinear dependence, being -0.841 as against -0.769. The character of the reaction of the yield to the degree of dryness

of the air during the period of ripening (the eighth period) likewise had a non-rectilinear character and is reproduced by the equation of the second order parabola:

$$y = 3.2384 - 0.3163H_g - 0.0773H_g^2; R = 0.528$$

Figure 2

Where H_g : the deviations from the mean long term value of $(E-e)$ at 1300 hours during the ripening period.

The graphic character of the dependence is reproduced by a curve of the following form (Figure 2).

This form indicates that marked saturation deficits during the ripening period, which are always accompanied at this time by high temperatures, are reflected very unfavorably in the yield of spring wheat. The action of this factor weakens as the dryness of the air declines, while at a considerably level atmospheric humidity (and at low temperatures) even a slightly negative effect of this factor is observed.

The correlation coefficient for this form of dependence is increased to -0.528 as against -0.389 under the hypothesis of a rectilinear connection between the fluctuations of the spring wheat yield and the saturation deficit. It is clear from all that has been set forth that the saturation deficit, which at first served as a combined index for the period from the fourth decade after the beginning of sowing to and including the ripening period, is now in this detailed analysis of the form of dependence, split up into two periods: the first from the fourth decade to and including blooming, and the second covering the ripening phase. Analysis of the

form of dependence of the yield on the number of days with rainfall 1 millimeter and over showed that a linear equation is entirely satisfactory for reflecting this dependence: raising the degree of the equation to the third degree leaves the correlation coefficient unchanged.

Thus in concluding our analysis of the south Trans-Volga, which at the present time is a part of Saratov oblast, we dwell on the following dominating factors that are responsible for the fluctuation in the yield of spring wheat on peasant farms from 1886 to 1915.

The number of rainy days from the beginning of the sowing down to the ripening occupies the first place. The form of the dependence is rectilinear. The correlation coefficient is 0.95. The regression equation is: $y_d = 35.0369 + 2.5969_1 D_7$.

The saturation deficit from the fourth decade after the beginning of sowing to and including the blooming, occupies the second place. The form of the dependence is curvilinear. The correlation coefficient is 0.841, and the regression equation is as follows:

$${}_1Y_H = -0.848610 - 0.150422_1 H_7 + 0.005901_1 H_7^2 - 0.004389_1 H_7^3.$$

The saturation deficit during the ripening period occupies the third place. The form of the dependence is curvilinear. The correlation coefficient is 0.528, and the regression equation is as follows: ${}_2Y_H = 3.2384 - 0.3163H_8 - 0.0773H_8^2$.

The values of y_d , ${}_1Y_H$, ${}_2Y_H$ were computed for each year by the equation. We then found the regression equation in which y (the deviation of the yield from the "normal" corresponding to

each year) is placed in dependence on the values of y_d , $1y_H$ and $2y_H$.

Thus the fluctuations in the yield of spring wheat also reflect the curvilinear dependence of the yield on the last two factors.

$$y = 0.6954y_d + 0.35441 \ 1y_H + 0.28327 \ 2y_H \quad R = 0.955.$$

For details see Appendix 2 to this chapter, and also diagram 6.

[see diagram 6 on following page]

This equation, however, contains a systematic error, which makes the correlation coefficient higher than it should be. Elimination of the error reduces the coefficient, but only very slightly, to 0.946, while this new numerical expression suffers from a systematic error on the other side -- it is very slightly lower than the true correlation coefficient. Thus the real correlation coefficient, denoted by the symbol ρ , lies within the limits: 0.946 0.955. (In making the connection we consider each stage as a separate variable. The connection thus proceeded with the following data: (1) number of observations = 30; (2) number of all variables (including y) = 7; (3) empirical correlation coefficients = 0.955.) Elimination of the systematic error in equation V leaves an almost unchanged correlation coefficient, reducing it from 0.932 to 0.930. (This is entirely natural: in this equation, with thirty observations, we have only three variables in all, and the empirical correlation coefficient is quite high.) Comparison of these two corrected correlation coefficients, 0.946 and 0.930, convinces us that it proved useful to take the curvilinearity into consideration, since the increase in the correlation coefficient is a reality, and not a fiction due to the increase in the number of variables.

DIAGRAM 6

YIELD OF SPRING WHEAT ON THE PEASANT FARMS OF THE TRANS-VOLGA OF THE LOWER VOLGA

_____ Observed yield

----- Yield calculated by the regression equation:

$$1y = 0.69541 y_d + 0.35441 1y_H + 0.28327 2y_H ; R = 0.955$$

Vertical line at left side of diagram: DEVIATION OF YIELD FROM NORMAL.

255

Thus the coefficient of multiple correlation taking into account the influence of all three factors included in the investigation as well as the curvilinear form of the dependence, is characterized by a value equal to 0.946, which is close to unity. This means that sixty-eight percent of the yield fluctuations are already accounted for, and that only thirty-two percent of "unsolved" deviations of the yield from the trend level of productivity now remain. Such deviations are attributable to all other factors (under the condition of their independence from those included in the investigation), or attributable to all the nonalignments caused by the approximate character of the equation and the inexactness of the materials. The indices of yield deviation calculated for each year by the resultant equation approach very closely to the actual yields, as a glance at the last two columns of the table in Appendix 2 will convince us. Of course complete coincidence is not observed, for this is possible only with a correlation coefficient equal to unity. In six cases out of thirty, however, a rather large divergence is observed between the actual yield indices and those calculated by the regression equation, namely in 1892, 1895, 1908, 1893, 1900, and 1905. These divergences ranged from 4.6 to 8.2 puds.

It would be possible to reduce the non-coincidence only by introducing new supplementary factors, but these would not even be third-order factors, but of even less influence on the yield, for the aggregate degree of their action cannot be higher than the correlation coefficient 0.32. (With a correlation coefficient of 0.946, the total influence of all the remaining factors (not correlated with those already used), is equal to a correlation coefficient of 0.32, and if they are included as supplementary factors in the investigation, leads to an aggregate correlation

coefficient equal to unity, that is to a functional dependence.) This exceedingly insignificant effectiveness of these factors makes it very difficult to segregate them from the "residual", "unsolved" divergences between the actual and the calculated indices. This question is still more complicated by the fact that the effectiveness of these low-order factors is apparently subject to change, depending on this or that action of the dominant factors. At least the following comparison of the non-alignments of the observed and calculated yields leads us to this thought. The years with the poorest yields give very small divergences: and the indices calculated by the regression equations coincide with the actual indices. The same is also observed for the most productive years, as may be seen from the following comparison:

YEARS	DEVIATION OF THE YIELDS (IN PUDS PER DESYATIN)	
	CALCULATED BY THE REGRESSION EQUATION	ACTUAL
1911	-23.0	-22.0
1906	-20.6	-20.5
1891	-17.3	-15.3
1894	+24.2	+25.0
1896	+19.5	+19.3
1913	+17.6	+15.4
1915	+16.8	+16.4
1904	+16.4	+17.1

The largest non-coincidences, both positive and negative, fall within the middle intervals of the computed indices, to which the "average" degree of dryness or humidity of the weather corresponds. All of this leads to the thought that the action of lower-order factors is not the same in different years. In years with an extreme combination of the dominant factors (rainlessness, high tem-

peratures, dryness of the air), the influence of the harmful supplementary factors is already unable to do any more damage than has already been done by the drought, while the positive influences are "cut off" by the same drought. In years with very favorable basic weather factors (abundant rain, cool weather, high atmospheric humidity), these factors overcome the harmful supplementary factors, while the positive factors can add nothing to what, at the given level of cultural practice, is supplied by exceptionally favorable weather. "Average" weather, however, affords more room for the play of the lower-order factors.

Thus on the southern steppes of the Trans-Volga, which at the present time form a part of Saratov oblast, the small, subdivided, and exploited present farm was completely delivered into the power of the elemental forces of the weather.

Even the insignificant uptrend of productivity could not prevent catastrophic crop failures. The crop failures of 1906 and 1911 took place during the last prerevolutionary decade.

Let us see how the estates reacted to the caprices of the weather. Their higher level of cultural practice was also responsible for their higher level of productivity. The average yield was also responsible for their higher level of productivity. The average yield of the peasant farms throughout the entire thirty-year period was 23.6 puds (sic!), while the estate yield was 31.5 puds per desyatin, or 7.8 puds (33 percent) more. But even this high level of productivity did not save the estates from the action of the elemental weather forces. The estates suffered the same fluctuations in the yields of spring wheat and were subject to the influence of the same weather conditions as the peasant farms. These fluctuations took place, however, about a higher level. (The same

thing had been noted already in the first chapter, devoted to rye.)

The correlation coefficient between the yield fluctuations and the number of rainy days from the first to the seventh period, inclusive, is $+0.904$ for the estates and $+0.915$ for the peasant farms. The accumulated saturation deficits from the fourth to the eighth period correlates with the estate yield to the extent of 0.807 and for the peasant yield to the extent of -0.814 . Finally, the regression equation for the fluctuation of estate yield, set up for these two factors, evaluates the aggregate influence of these two factors by a correlation coefficient of 0.924 (0.932 for the peasant farm).

In years of serious crop failures (for example, 1891, 1906, 1911) the estates also gave very low harvests, which were not very considerably different from those of the peasant farms, as may be seen from the following Table 4:

[see Table 4 on following page]

The first group of years, with very severe drought, gave harvests on the peasant farms slightly in excess of the amount of seed used. The estate harvests were better, but even they gave only twelve puds per desyatin, which was lower than the average yield obtained by the peasants in the second group of years, which though they were also years of drought, still had a higher rainfall (the mean was 44 millimeters against 29 millimeters for the first group). Things are different in the rest of the table. The superior cultural practice on the estates led to their obtaining under appreciably inferior meteorological conditions the same yield of spring wheat, or even a little more than that, as the peasant yield harvested in the group of years with the best meteo-

TABLE 4
NOVOUZENSK AND NIKOLAYEVSK UYEZDY

GROUPS OF YEARS BY AMOUNT OF RAINFALL IN FIRST SEVEN DECADES AFTER BEGINNING OF SOWING	MEAN RAINFALL	MEAN YIELD OF SPRING WHEAT IN PUDS PER DESYATIN	
		PEASANT FARMS	ESTATES
Group 1: 5 years with lowest rainfall (14 - 39 mm.)	29 mm.	7.9	11.9
Group 2: 5 years with rainfall 40 - 48 mm.	44 mm.	14.9	21.9
Group 3: 5 years with rainfall 49 - 55 mm.	52 mm.	21.8	28.4
Group 4: 5 years with rainfall 56 - 71 mm.	62 mm.	25.4	35.2
Group 5: 5 years with rainfall 72 - 86 mm.	80 mm.	32.6	41.5
Group 6: 5 years with rainfall 96 - 110 mm.	103 mm.	39.6	50.1

Note: The peasant and estate yields of adjoining groups but comparable values are joined by the sign

186

rological conditions.

For instance, during the fourth group of years, with mean rainfall of sixty-two millimeters for the seven decades, the estates obtained a yield somewhat higher than the peasant yield in the group of years with mean rainfall of eighty millimeters. This same group of years gave a higher estate yield of winter wheat than the peasant yield in the years of most abundant rainfall for the region under study, amounting to one hundred and three millimeters on the average. Exactly the same conclusions result from Table 5 which gives the data for the two former northern uyezdy along the Volga, Stavropol' and Samara. These two uyezdy were characterized by more favorable climatic conditions than Novouzensk and Nikolayevsk, and in general had the highest productivity of all the seven uyezdy of Samara guberniya. Since the accumulated rainfall for eight decades from the beginning of sowing proved more effective for the former Stavropol and the former Samara uyezdy than that for seven decades, this grouping of years is based on the rainfall for eight decades instead of seven.

[see Table 5 on following page]

The conclusions just set forth on the meteorological condition and on the relation between peasant and estate yields of spring wheat in the former Samara guberniya coincide completely with the conclusions set forth in Chapter I with respect to the yield of winter rye in the same guberniya. The conclusions on the variability of the spring wheat yield on the peasant farms and estates are identical in exactly the same way: the relative variability of the estate yields is somewhat lower than that of the peasant yields, while the absolute variability (in puds per desyatina) is a little higher. This is clear from a comparison of the extreme groups from

TABLE 5
SAMARA AND STAVROPOL' UYEZDY

GROUPS OF YEARS BY AMOUNT OF RAINFALL IN FIRST EIGHT DECADES AFTER BEGINNING OF SOWING	MEAN RAINFALL	MEAN YIELD OF SPRING WHEAT IN PUDS PER DESYATIN	
		PEASANT FARMS	ESTATES
Group 1: 5 years with lowest rainfall (41 - 54 mm.)	49 mm.	9.8	13.8
Group 2: 5 years with rainfall 77 - 83 mm.	81 mm.	31.2	38.3
Group 3: 5 years with rainfall 84 - 91 mm.	86 mm.	34.2	40.3
Group 4: 5 years with rainfall 92 - 106 mm.	99 mm.	37.5	50.3
Group 5: 5 years with rainfall 109 - 119 mm.	113 mm.	44.1	52.1
Group 6: 5 years with rainfall 120 - 153 mm.	131 mm.	48.9	57.4

the last two tables:

FORMER	HOW MANY TIMES THE PRO-	DIFFERENCE BETWEEN THE
UYEZDY	DUCTIVITY OF THE SIXTH	PRODUCTIVITY OF THE SIXTH
	GROUP OF YEARS EXCEEDED	GROUP OF YEARS AND THAT
	THAT OF THE FIRST GROUP	OF THE FIRST GROUP, IN
		PUDS PER DESYATIN

	Peasant Farms	Estates	Peasant Farms	Estates
Southern:				
NOVOUZENSK	5.0	4.2	31.6	38.2
NIKOLAYEVSK				
Northern:				
SAMARA	4.9	4.3	39.1	43.6
STAVROPOL'				

We pass now to the analysis of the dependence of the fluctuations in the yield of spring wheat ("Poltavka" variety) at the Bezenchuk experiment station, which is located in the south of the former Samara uyezd from 1905 to 1929. The twenty-five year period from 1905 to 1929 inclusive, thus embraces not only prerevolutionary years but also years of Soviet power. This period includes 1921, a year of catastrophic drought and crop failure. (A total of six millimeters of rainfall was registered at Bezenchuk meteorological station in 1921 from the third April decade to the beginning of July, or seven decades in all (absolute rainlessness)). The Bezenchuk experiment station stood on a higher level of cultural practice than either the peasant prerevolutionary farm or the estate. This higher level of agronomic culture assured a higher level of productivity, which notably exceeded both prerevolutionary peasant

and estate yields. The average yield for the twenty-five years at Bezenchuk station was 61.8 puds, which was seventy-one percent higher than the prerevolutionary peasant productivity up to 1915.

In spite of this, the fluctuations of the spring wheat yield on the fields of the Bezenchuk agricultural stations were determined to no less an extent by the weather conditions and were dependent on the very same meteorological factors that influenced the fluctuations of the yield on the prerevolutionary peasant farms and estates.

Since the question of the effectiveness of the meteorological factors with respect to the fluctuation of yields at the Bezenchuk station has been discussed by me in detail in the work "The Dependence of the Yield Fluctuations of Winter and Spring Wheat in the Chernozem Belt" (from materials of the experiment stations) I shall not dwell in detail on this question and shall present only the resultant data of that work.

1. The rainfall from the third April decade to and including the last June decade had the greatest importance for the yield of spring wheat at the Bezenchuk station. The correlation coefficient between the yield of spring wheat and the total rainfall during this period is 0.925. The following table demonstrates the immense importance of this period.

[see Table 6 on following page]

The regression equation in which the yield of spring wheat on the fields of the Bezenchuk experiment station is determined, in tsentner per hectare, in dependence on the rainfall during the seven decades mentioned, is of the following form:

$$y = 1.76 + 0.08619 P.$$

TABLE 6

YEARS GROUPED BY AMOUNT OF RAINFALL FROM THIRD APRIL DECADE TO END OF JUNE	MEAN AMOUNT OF RAINFALL	MEAN YIELD			
		OBSERVED		CALCULATED FROM REGRESSION EQUATION	
		IN PUDS PER DESYATIN	IN TSENTNER PER HECTARE	IN PUDS PER DESYATIN	IN TSENTNER PER HECTARE
5 years with lowest rainfall: 6 - 38 mm.	23.6	23.5	3.52	25.3	3.79
5 years with rainfall 50 - 86 mm.	68.8	53.0	7.95	51.3	7.69
5 years with rainfall 88 - 97 mm.	93.4	67.2	10.08	65.4	9.81
5 years with rainfall 103 - 113 mm.	109.0	75.8	11.36	74.4	11.15
5 years with rainfall 115 - 184 mm.	140.1	89.0	13.34	92.3	13.83

The indices of yield by groups of years were computed by this regression equation. As may be seen from the preceeding Table 6, these computed indices differ very little from the factual data.

2. The final regression equation takes account of the influence of the rainfall and the dryness of the air. The latter is characterized by the readings of the evaporation power by the Vil'd evaporimeter. (If an investigation includes the indices of the evaporation power after Vil'd it becomes superfluous to use the saturation deficit ($E - e$), the more so since the indices of the Vil'd evaporimeter also reflect the influence of the force of the wind. These indices thus take account of three meteorological elements: (1) temperature, (2) atmospheric humidity, and (3) the force of the wind.)

The evaporation power from the third May decade to the third June decade inclusive exerts the greatest influence on the yield of spring wheat. The regression equation is as follows:

$$y = 6.1025 + 0.09194P_1 + 0.05671P_2 - 0.01434E_2; \quad R = 0.950$$

where P_1 is the rainfall from the third April decade to the second May decade, P_2 the rainfall of the last May decade and all of June, and E_2 the accumulated evaporation after Vil'd during the same interval. It is clear that these three factors are dominating, and that the primacy should be assigned to the rainfall. The activity of the other two factors is joined, about a month after the start of sowing, by the negative influence of atmospheric aridity. In a word, in spite of the considerably higher level of spring wheat productivity at Bezenchuk station (determined by superior cultural practice) the same factors of spring-summer weather, with no lesser force, influence the fluctuation of the yield

266

as acted on the prerevolutionary peasant farms and estates. (If the accumulated evaporation, according to Vil'd, for the first period, is also included in the equation, it becomes:

$$y = 0.0998 P_1 + 0.0225 P_2 + 0.0235 E_1 - 0.0202 E_2$$

R here is hardly increased: it is 0.956, while the plus sign before E_2 conceals the weak positive action of the temperature during the initial period in the life of spring wheat.)

[see Diagram 7 on following page]

The data on the vegetative periods at Bezenchuk cover the seventeen years from 1913 to 1929. The average date sowing began was 23 April; of inception of tillering, 16 May; of inception of blooming, 23 June; of ripening, 4 July; and of harvest commencement, 23 July. Thus the most effective period for the influence of meteorological factors covers the time from beginning of sowing to beginning of ripening.

Adjustment of the tabulation of the meteorological factors to the vegetative periods for seventeen years gives even higher correlation coefficients: yield at Bezenchuk station correlates with rainfall from start of sowing to start of blooming to the extent of 0.973.

It may not be superfluous to recall that this correlation coefficient speaks not only for the influence of the rainfall but also to a considerable extent of the action on the yield of spring wheat from the fields of the Bezenchuk station of other meteorological factors intercorrelated with the rainfall (high temperatures, saturation deficits).

There are still two more works on the Bezenchuk experiment station.

DIAGRAM 7

YIELD OF "POLTAVKA" SPRING WHEAT AT BEZENCHUK AGRICULTURAL EXPERIMENT STATION

_____ Observed yield

----- Yield calculated by regression equation:

$$Y = 6.1025 + 0.09194P_1 + 0.56705671P_2 - 0.01434 E_2 ;$$

$$R_2 = 0.950$$

840

(1) The published work of Professor D.I. Shchukin, entitled "Atmospheric Precipitation as a Factor in Productivity" and (2) the unprinted work of Ye. Kushevitskaya "Harmful Meteorological Elements and the Yield of Spring Wheat in the Middle Povolzh'ye." Professor Shchukin's investigation establishes the strong influence exerted by the May and June rainfall on the yield of spring wheat at Bezenchuk station (correlation coefficient ≈ 0.88). Kushevitskaya's work notes the inverse dependence of the spring wheat yield (a) on the number of days with harmful temperatures; (b) on the number of days with particularly high evaporation power (by the Vil'd evaporimeter); and (c) on the number of days with low relative humidity. The corresponding correlation coefficients are -0.78 , -0.81 and -0.68 respectively, while the coefficient of multiple correlation for the combined action of all three factors is 0.91 .

Thus the first work analyzes only the influence of soil drought [as in text; but apparently inconsistent with the preceding paragraph] on the yield, while the second work only covers the action of atmospheric drought. My own investigation takes account simultaneously of the influence of both elements. The conclusions of all three works are in complete agreement. (In my work I established that the most effective period for the influence of rainfall covers the seven decades from the third in April to the last in June inclusive. And, in fact, exclusion of the third April decade reduces the correlation coefficient from 0.925 to 0.882 , while addition of the second April decade likewise reduces it to 0.915 .)

Let us see how high the variability was in the spring wheat

yield at Bezenchuk station. The data presented above make it plain that it was very considerable. Under the conditions of the absolute drought in 1921, when the total rainfall from the third April decade to the end of June amounted to six millimeters, 1.3 tsentner per hectare were obtained, while in 1915, with one hundred and eighty-four millimeters of rainfall during the same period, the Bezenchuk station yield of spring wheat reached the record figure of 17.7 tsentner. The absolute variability, expressed as the standard deviation, was 3.859 tsentner per hectare, while the relative variability is expressed by the figure 41.7 percent.

The absolute variability for the prewar peasant farm from 1886 to 1915 was 13.07 puds per desyatin, and the relative variability was 48.9 percent.

Thus the absolute variability of the spring wheat yield at Bezenchuk station was considerably higher, while the relative variability was somewhat lower. This comparison of the stability of the spring wheat yield is not, however, irreproachable: the Bezenchuk data relate to a single farm, to a single producing unit, while the data on the peasant yields characterize the totality, the collectivity of peasant farms, and a multiplicity of productive units distributed over the territory of the former Samara guberniya. By virtue of the single circumstance that the meteorological factors, mainly the rainfall, are more constant over a large territory than at a single point of this territory, the stability of the yield (other things being equal) of a collectivity of producing units spread over a wide territory should be greater than that of a single farm, a single producing unit.

If the cultural practice at Bezenchuk station had been pre-

valent in all the farms distributed over the territory of the former Samara guberniya, then, given the same dependence of the spring wheat yield on the meteorological factors (reflected by the regression equations presented above), the stability in the yield of this collectivity of farms using the "Bezenchuk" cultural practices would have been higher than that of "Bezenchuk alone."

The comparison between the stability in the yield of a collective and that of a single producing unit is thus illicit.

Let us try to overcome this difficulty and render the stability of prerevolutionary peasant yields comparable with that of the Bezenchuk station yields of spring wheat.

As we have seen, the rainfall from the third April decade to the end of June, inclusive, was a factor that influenced the fluctuations of these yields at Bezenchuk station. This same rainfall also influenced these yields on the peasant farms, as is plain from the following table.

[see Table 7a on following page]

Let us now explain the data in this table. The first column shows the range of the rainfall in each group of years. The thirty-year period from 1886 to 1915 is divided into five equal groups (pentiles) according to amount of rainfall. The second column gives the mean rainfall for each of these five groups. The third column gives the mean deviations about the mean productivity for each year. (We have already observed more than once that the uptrend in productivity forces us to segregate the trend component by the method of least squares and to compute the deviations of the yield about this trend line.) The fourth column requires de-

TABLE 7a

YEARS GROUPED BY RAINFALL DURING THE THIRD APRIL DECADE, MAY AND JUNE, IN THE FORMER SAMARA GUBERNIYA (9 STATIONS)	AVERAGE RAINFALL	DEVIATIONS OF PEASANT YIELDS IN PUDS PER DESYATIN	YIELD IN PUDS PER DESYATIN AT HIGHEST NORM OF PRODUCTIVITY ATTAINED ON PEASANT FARMS UP TO 1915
6 years with lowest rainfall: 43.7 - 57.7 mm.	50.8	-16.8	19.5
6 years with rainfall of 65.8 - 78.5 mm.	71.3	- 4.2	32.1
6 years with rainfall of 78.5 - 89.0 mm.	83.0	- 1.6	34.7
6 years with rainfall of 97.0 - 114.2 mm.	108.2	' 9.2	45.5
6 years with highest rainfall: 118.9 - 140.8 mm.	129.0	'13.4	49.7

tailed explanation. This column gives the indices of the absolute yield on the assumption that the level about which the yields fluctuate is that attained by the peasant farm up to 1915, i.e., that of the maximum attainments of the prerevolutionary peasant farm. This level of peasant productivity in 1915, according to the data of the former Central Statistical Committee, is determined by the method of least squares to be 31.9 puds per desyatin. It has however been repeatedly noted by various authors (of whom the first was Professor Fortunatov in 1893) that the crop data of that Committee were somewhat underestimated, and therefore these data require correction. The correction to these data for the spring wheat yields in the former Samara guberniya was determined to be 13.9 percent. The maximum level attained by the peasant farm up to 1915 is thus expressed by the index 36.3 puds per desyatin (31.9×1.139), or 5.44 tsentner per hectare. With this level as a starting point, the absolute yields are readily calculated from the deviations from the mean productivity given in column three.

For instance, the first group, consisting of the most arid years, gives a deviation of -16.8 from that level. Consequently the yield, expressed in absolute terms of harvest in puds per desyatin, is $36.3 - 16.8 = 19.5$ puds. The yield of the last group, by analogy to this, is $36.3 + 13.4 = 49.7$ puds.

Thus the data in column four of Table 7a represent the harvests in puds per desyatin, by groups of years with differing amounts of rainfall, on the assumption that the yield of spring wheat on the territory of the former Samara guberniya throughout the entire thirty-year period fluctuated about the maximum level of productivity (It is plain that as many schemes can be set up as there are

years in our period, i.e., thirty schemes (our period is 1886 to 1915). For instance, the absolute values of the productivity according to the year-groups of Table 7a could be also computed on the basis of the minimum level of productivity of the prewar peasant farm; for this purpose we would have to use as a base the mean productivity corresponding to the initial year of the period taken (1886). But of all these schemes, there is only one that can compete with the basic scheme of Table 7a. This is the one based on the mean productivity corresponding to 1910, and this is why.

There are eleven years, from 1905 to 1915, for which we have information on both peasant yields in the former Samara guberniya and the Bezenchuk station yields. Therefore in comparing the productivity we should use either the maximum productivity ever attained by the prerevolutionary peasant farm, or the midpoint of the mean annual productivities over the eleven-year period. The mean productivity corresponding to the middle or median year 1910, constitutes this midpoint of the means.

The first variant, is undoubtedly the most correct, since a comparison of the productivities according to the groups of Table 7a should take due account of Bezenchuk's superiority, without at all exaggerating that superiority. This superiority is somewhat exaggerated in the second variant, since the "achievements" of the peasant farms from 1910 to 1915 are included in the productivity difference.

We note that the second variant differs only very slightly from the principal one. The mean productivity corresponding to 1910

is 30.3 puds. The correction for the underestimation by the data of the TsSK increases this index to 34.5 puds per desyatina. Consequently, the index of productivity for the first group of years is lowered from 19.5 to 17.7 puds; it is 30.3 puds for the second group, 32.9 puds for the third, 43.7 puds for the fourth, and finally, for the fifth and highest group, it is 47.9 instead of 49.7 puds in the first and principal variant. In the first variant the productivity of the last and highest group is 2.55 times that of the first group, while in the second variant it is 2.71 times higher.) against which the prerevolutionary peasant farm "knocked" (the 1915 level).

Let us assume that, alongside of the prerevolutionary peasant farm with its primitive cultural practice, there also existed in Samara guberniya, during the period under study (1886-1915), and existed in sufficient numbers (say for instance that there were as many as the estates), progressive productive units employing the cultural practices of the Bezenchuk experiment station. We put the question as to what the productivity of these progressive farms would have been in the groups of years, from 1885 to 1915, with varying amounts of seven-decade rainfall (during the last April decade, all of May, and all of June). We know that the rainfall during this interval had a very strong influence on the spring wheat yield at Bezenchuk station: the correlation coefficient is 0.925, and the regression equation that determines that yield in dependence on the rainfall is: $y = 1.76 + 0.08619 P$.

We see from Table 7a that the peasant yield is computed at 19.5 puds for the six years of worst drought, with mean rainfall of 50.8 millimeters for the seven decades. By the use of the equation

just given, we determine the yield to be expected in each of these six years for a "Bezenchuk" type farm, according to the rainfall in each year, and then find from these indices the mean value for the six-year period. It is 6.14 tsentner per hectare. We proceed similarly for the remaining groups, (Since the dependence of the fluctuations of spring wheat at Bezenchuk station on the rainfall is well reproduced by a linear regression equation, the complicated procedure of computing the productivity indices by groups, for Bezenchuk type farms, may be replaced by substitution of the mean rainfall indices in the regression equation. Thus, for instance, for the first group of years, we substitute the value $P = 50.8$ in the regression equation, for the second group of years we take P equal to 71.3, etc. We see from the example of Table 6 that the five-year means according to the regression equation come very close to the means of the actual Bezenchuk yields. As we have already remarked, the Bezenchuk station is situated in the steppe portion of the former Samara uyezd, near the boundary line of the former Nikolayevsk uyezd. To the south of that station lies a territory with worse climatic conditions than it enjoys, while not far to the north commences the wooded steppe, where climatic conditions are milder. Thus the climatic position of Bezenchuk is very similar to the average climatic conditions of the former Samara guberniya, which statement is confirmed by the following data: the mean twenty-five year rainfall for the seven decades indicated is 87.0 millimeters at Bezenchuk and 88.5 millimeters for the former Samara guberniya. The comparison of productivity is thus entirely in order.) and as a result obtain the indices given in Table 7b.

[see Table 7b on following page]

GROUPING OF YEARS BY AMOUNT OF RAINFALL	TABLE 7b		"BEZENCHUK" TYPE FARM YIELDS	
	MEAN RAINFALL (millimeters)	PEASANT YIELDS (in puds per desyatin)	(in puds per desyatin)	(in tsentner per hectare)
Group 1: 6 years of worst drought	50.8	19.5	41.0	6.14
Group 2: 6 years	71.3	32.1	52.8	7.91
Group 3: 6 years	83.0	34.7	59.4	8.91
Group 4: 6 years	108.2	45.5	73.4	11.09
Group 5: 6 years of best rainfall	129.0	49.7	85.9	12.68

We see from the indices of this table that: (1) the difference between the absolute prerevolutionary peasant productivity during the first group of the worst drought years and that during the last group of the moistest years was 30.2 puds per desyatin, while under the conditions of the "Bezenchuk" cultural practice it would have been 44.9 puds, i.e., one and one-half times as great; (2) but the relative difference, on the other hand, is appreciably lower: the most humid group of years gives a yield under Bezenchuk cultural practice which is about double that of the most arid years. The corresponding increase for the peasant farms is about two and one-half times as great (the more precise coefficients are 2.09 and 2.55 respectively); (3) in spite of the fact that the yield fluctuations under "Bezenchuk" cultural practice were determined by the same spring-summer weather conditions that also affected the fluctuation of the spring-wheat yield on the prerevolutionary peasant farm (with the effectiveness of the meteorological factors not being weaker by any means), the considerably superior "Bezenchuk" cultural practice would have resulted in the years of worst drought (Group 1) giving a yield that was even higher than the peasant yield under average rainfall conditions: 41.0 against 34.7; while the second group of years, with rainfall definitely under average, would have given a yield of spring wheat better than the peasant yield under the best rainfall conditions: 52.8 puds against 49.7 puds. (On comparing Tables 6 and 7b we see that grouping the years by rainfall at Bezenchuk stations results in a higher variability of yield obtained on Bezenchuk agronomic principles than is the case when the years are grouped by the rainfall throughout the entire territory of the former Samara guberniya. The reason for this difference, as can be seen from these tables, is the variability

lity of rainfall over a large area is less than that at a single point of that area. The rainfall indices for the whole of the former Samara guberniya have been determined by me on the basis of the observations of nine stations that have been in uninterrupted operation for thirty years. The standard deviation of these indices is 28.2 millimeters, while at Bezenchuk station the standard deviation is 41.4 millimeters, or almost one and one-half times as great.

We have thus established the immense influence of the spring-summer weather on the fluctuations in the yield of spring wheat in the former Samara guberniya, at three levels of cultural practice: (a) at the lowest level, that of the prerevolutionary peasant farm; (b) at a higher level, that of the prerevolutionary estate; and (c) at the highest of these three levels, that of the Bezenchuk experiment station. At the same time we have shown the great practical value of cultural practice, especially in comparing the yield of spring wheat from peasant farms with that obtained by the Bezenchuk station, while this comparison was carried out not only with respect to mean long-term yields but also, which is particularly important, by groups of years with varying rainfall conditions.

Let us now proceed to discuss the results of the studies made on the yield of spring wheat on the territory of the former Saratov guberniya, and the yield of the Saratov agricultural station located thereon.

Two works have been devoted to the elucidation of the influence of meteorological factors on the fluctuations of the spring wheat yield on the peasant farm. We refer to the work of A.M.

Protasov, already cited by us above, entitled "The Dependence Between the Yield of Rye and Wheat and the Meteorological Conditions in Certain Uyezds of the Lower Povolzh'ye", and the work of N.P. Kuritsyn "On the Question of the Productivity of Wheat" (Leningrad, 1930). Both of these works note the strong influence of the spring-summer weather conditions on the yield of spring wheat on the peasant farms of the former Saratov guberniya. In Protasov's work the net significance, for the guberniya as a whole, of the positive influence of the rainfall and the negative influence of high temperatures and saturation deficits, during the May-June period, is expressed by a coefficient of correlation equal to 0.79. He points out that, in the northwestern uyezds, which are marked by a moister climate, this effect of rainfall and temperature is appreciably weaker. The uyezds on the Volga have correlation coefficients ranging from 0.76 to 0.88. Kuritsyn, using the method of calendar decades, considers the most effective period is that from the second May decade to the first June decade inclusive. The coefficient of multiple correlation, expressing the influence of both the positive influence of rainfall and the harmful action of high temperature during this period, reaches 0.84. (These coefficients are high, but lower than those established by me for the peasant farms of the former Samara guberniya. On the one hand, this is explained by the fact that the northwestern uyezds, which have a milder climate, are included in the studies, made on a guberniya-wide scale. This circumstance, however, is evidently not of great importance, since large amounts of wheat were not sown in the Northwest of the former Samara guberniya. On the other hand it must be borne in mind that the indices of yield and meteorological conditions are more carefully worked up in my works. Neither in

the work of Protasov nor in that of Kuritsyn is the investigation conformed to the vegetative periods, as I did for the South of the former Samara guberniya. I employ only the data from stations with an unbroken period of operation. The yield series in Protasov's work is not above reproach, and he operates exclusively with monthly indices of meteorological observations. Kuritsyn uses the ten-day summaries of the "Monthly Bulletin" of the Central Geophysical Observatory (GCO), with a small number of stations, selected in a rather random and inconstant manner.)

Neither work touches the question of the influence of meteorological factors on the fluctuations of the yield from the estates, which stood at a higher level of cultural practice than the prerevolutionary peasant farms. For this reason the slight additions made by me to the works of these authors will not be accounted superfluous.

In Table 8, prepared by me, the years are grouped by the mean temperature (at 1300 hours) for the second and third months following the first decade with a temperature of five degrees and higher. This grouping has already been used in analyzing the fluctuations of the rye yield. In each group the mean deviations of the spring wheat yield from the trend level are computed separately for the prerevolutionary peasant farms and estates. The yields have then been calculated in absolute terms on the assumption that their fluctuations were based on maximum level of cultural practice attained in 1915, and were about the mean yields corresponding to that level, which was 31.6 puds per desyatin for the peasant farm and 45.8 puds per desyatin for the estate. The yield data have been adjusted by correction of 12.5 percent to com-

compensate for underestimation by the former Central Statistical Committee.

[see Table 8 on following page]

It is clear from this table that both peasant and estate yields of spring wheat reacted strongly to the spring-summer temperatures, in spite of the difference in the cultural practices of these farm-types. The higher level of cultural practice on the estates led to a very appreciable practical result: under worse conditions they obtained the same yield harvested by the peasants under distinctly better temperature conditions. Thus, for example, the group of the hottest years gave a harvest on the estates that was even somewhat higher than that on the peasant farms under average temperature conditions, while the second group of years, with unfavorable temperature conditions gave the estates the same yield harvested by the peasants under rather favorable temperature conditions. In a word, we have here a new confirmation of one of the fundamental theses of our work. We see in exactly the same way, in complete agreement with the conclusions previously drawn by us, that the absolute variability of the estate yield almost coincides with the absolute variability of the peasant yield (somewhat exceeding the latter): the range between the peasant yields of the extreme groups is 27.3 puds, while it is 29.5 puds between the estate yields of the same groups. The relative variability, however, is lower for the estate yield. The peasant productivity in the last group of years is 2.31 times that in the first group of years, while for the estates it is 1.88 times greater. The standard deviation of the peasant yield over the thirty-three year period is 11.6 puds, that of the estate yield, 13.8 puds, while the relative variability is 43.2 percent and thirty-eight and

TABLE 8

YEARS GROUPED BY TEMPERATURE AT 1300 HOURS FOR SECOND AND THIRD MONTHS AFTER THE FIRST DECADE WITH TEM- PERATURE FIVE DEGREES AND OVER	DEVIATIONS OF YIELDS FROM LEVELS OF PRODUCTIVITY CORRESPONDING TO EACH YEAR		YIELD IN PUDS PER DESYATIN, COMPUTED ON BASIS OF PRODUC- TIVITY LEVEL ATTAINED IN 1915	
	Peasant Farms	Estates	Peasant Farms	Estates
6 hottest years: 25.1 - 26.1 degrees	-10.7 puds	-12.3	20.9	33.5
6 years with temperature 23.4 - 25.1 degrees	- 7.4	- 8.5	24.2	37.3
6 years with temperature 22.7 - 23.2 degrees	- 3.6	- 0.3	28.0	45.5
6 years with temperature 21.4 - 22.5 degrees	5.1	3.8	36.7	49.6
6 coolest years: 18.5 - 20.8 degrees	16.6	17.2	48.2	63.0

one tenth percent respectively. (We indicate the second variant of the comparison between the peasant farms and the estates. This variant is based on the mean productivity over the thirty-year period, shown separately for the peasant farms and the estates:

	PEASANTS	ESTATES
1st group	19.0	28.5
2nd group	22.3	32.4
3rd group	26.1	40.5
4th group	34.8	44.7
5th group	46.3	57.9

Finally we present a comparison of the yields, without taking the trend into account at all, and without correction for underestimation by the TsSK:

	PEASANTS	ESTATES
1st group	16.7	26.1
2nd group	20.0	29.5
3rd group	23.1	35.2
4th group	30.8	38.2
5th group	41.5	52.5

Both of these groupings, in general, confirm the basic thought.)

Let us now examine the nature of the dependence of the fluctuations in the spring wheat yield under the higher cultural practice of the Saratov experiment station. Two published works are devoted to the analysis of this dependence: N.F. Yemel'yanov "Variety Testing of Spring Wheat from 1915 to 1925", and B. Shneydyuller "Critical Periods and the Connection Between the Yield of Winter and Spring Wheat and Meteorological Factors under the Conditions of Dry Farming and Irrigated Farming." The Yemel'yanov work uses data for ten years, while that of Shneydyuller includes a longer

series: material covering the seventeen years from 1915 to 1931 inclusive. We therefore dwell on the latter work, without, however, deeming the reservation superfluous that there are no contradictions between the two. In addition I also prepared supplementary material dealing mainly with the influence of meteorological factors on the spring wheat yield where snow-retention was practiced.

On plots without snow-retention the eighteen-year mean yield at the Saratov experiment station was 10.91 tsentner per hectare, or 72.8 puds per desyatin. This station thus notably surpassed the achievement of the Bezenchuk station, which had a mean yield of 61.8 puds per desyatin. The Saratov yield, without snow-retention, was more than double the peasant yield of 31.6 puds reached by 1915 (more precisely, it was 2.28 times as high), and was sixty percent higher than the estate yield of 45.8 puds.

Shneydermyuller in his work establishes the following indices of spring wheat yield in dependence on the weather conditions during the various vegetative periods.

[see Table 9 on following page]

"A day is considered humid if the evaporation power, computed by the formula $0.5 (E - e)$, does not exceed the amount of rain falling or remaining from previous days." (Shneydermyuller, *op. cit.*, p. 8). In this formula $(E - e)$ is the saturation deficit. Thus whether any given day is accounted dry or humid depends in the last analysis on the use of the Mayer formula for computing the water supplies in reservoirs from free-water surfaces: $P \frac{E - e}{2}$, where P is the amount of rainfall on one square meter of surface, and $\frac{E - e}{2}$ is the daily evaporation of the water. It can be readily seen that Shney-

TABLE 9
CORRELATION COEFFICIENTS OF THE SPRING WHEAT YIELD AT SARATOV EXPERIMENT
STATION WITH THE NUMBER OF HUMID DAYS AND THE
RAINFALL

VEGETATIVE PERIODS	WITH NUMBER OF HUMID DAYS	WITH AMOUNT OF RAINFALL
From sowing to germination	0.215	no data
Germination to tillering	0.594	no data
Tillering to heading	0.847	0.775 0.818
Heading to maturation	0.440	no data

from germination to onset of maturation (my supplement to this work)

Shneydyuller (as well as Yemel'yanov) is really taking combined account as well, by his index of the number of humid days, or rainfall, temperature (for E is a function of the temperature) and of the atmospheric humidity. (I am not a partisan of such a composite use of both rainfall and evaporation in a single formula, for the analogy between changes in water stocks of reservoirs and soil moisture stocks is incomplete. And, in fact, I have shown, in my unpublished work "The Influence of Meteorological Factors on the Variation of Soil Moisture Stocks", that the coefficient of $(E - e)$ does not remain constant at 0.5. With heavy rainfall it is higher than 0.5, and with scanty rainfall it is lower. As may be seen from all the foregoing, I prefer to use $(E - e)$ as a separate variable in the regression equation.)

This combination of the effectiveness of soil and atmospheric drought into a single index increases the correlation coefficient to 0.847 from its value of 0.775 for the rainfall alone. Shneydyuller has unfortunately not computed the correlation coefficient between yield and rainfall for the other vegetative periods, and it is therefore hard to form any opinion on the usefulness of this combined use of rainfall and evaporation power $(E - e)$ in a single formula. My supplement showed that the rainfall had a positive influence on the spring wheat yield at Saratov experiment station almost throughout the whole life of that plant, from germination to the onset of maturation: the accumulated rainfall during this period correlates with yield to the extent of 0.818.

Shneydyuller includes a dry-wind index to establish a closer dependence of yield on the meteorological factors during the most effective period (tillering to heading). This index for the

separate periods in different years is taken as the accumulated wind velocities at 1300 hours on dry days. Inclusion of this factor brings the correlation coefficient for the Saratov experiment station up to 0.94. (We remind the reader that by including the Vil'd evaporimeter readings as an index of atmospheric drought in the Bezuchuk station study, we also took account of the influence of the wind.)

The rainfall from tillering to heading, the saturation deficit during the same period, and the amount of dry wind are thus the principal factors in the yield fluctuations at Saratov experiment station. But the meteorological factors in the other vegetative periods as well likewise affected the yield of spring wheat. My supplement to the Shneydmyuller work shows that the rainfall from germination to the onset of tillering supplemented the action of the rainfall from tillering to heading, just as did the rainfall after heading down to the onset of maturation.

We also note the interesting dependence established by Yemel'yanov between the weight of 1000 kernels of grain and the temperature conditions from heading to maturation inclusive. If the period from tillering to heading was humid, the temperature from heading to maturation inclusive had a positive influence on kernel weight; if this foregoing period was dry, however, the influence of temperature was negative.

Shneydmyuller analogously traced the influence of the meteorological factors during the heading phase and until and including maturation, in dependence on the meteorological conditions in the preceding period from tillering to maturation inclusive. In dry years the connection between the yield and the number of

humid days of the concluding period was positive, with a correlation coefficient of 0.36, but following humid periods from tillering to maturation inclusive, this connection was weakly negative, with correlation coefficient of -0.1215.

Thus in spite of the higher level of cultural practice at Saratov experiment station than on the peasant farms and the estates, the spring wheat yield still fluctuated about its higher level of productivity under the influence of the same elemental forces of nature that determined the fluctuations of the yield on the prerevolutionary farms.

Unfortunately I had no opportunity to prepare a table analogous to Table 7b, which revealed the practical value of the higher cultural practice at Bezenchuk experiment station as compared to the low level of cultural practice on the prerevolutionary peasant farm. But even without this study of the material, it may be said in advance that the effectiveness of the cultural practice at the Saratov station was far higher, since its achievements were very appreciably greater than those of the Bezenchuk station. Indeed, we see from Table 8 that the estates gave about the same yield in the group of hottest years that the peasant farms obtained at average temperature conditions, while the estate productivity was forty-five percent higher than that of the peasants (45.8 pud against 31.6). The productivity at Bezenchuk station was seventy-one percent higher than the peasant productivity. Accordingly, not only did the first group, with the years of the worst drought, give a yield with the Bezenchuk cultural practice equal to the peasant yield under average meteorological conditions, but the second group of years, with reduced rainfall, assured a yield, under Bezenchuk cultural practice, higher than the

peasant yield in the group of the most humid years. The productivity of the Saratov experiment station exceeded the peasant productivity by even more: it was 2.30 times the latter. We are therefore able to maintain that the group of the driest years yielded at least as much, with the cultural practice of this station, as the peasant farms under better than average meteorological conditions, and equal to the estate yield under average meteorological conditions (its productivity was an average of sixty per cent above that of the estates); and that the third group, with average meteorological conditions, exceeded the productivity attained by the estates in the best years of all, as far as weather was concerned. As a general rule, on the fields of the experiment stations, spring wheat was included in a definite crop rotation, autumn plowing was done to a depth of eighteen to twenty centimeters, with a preliminary removal of the stubble; and early sowing done with a row drill, and the plowland cultivated before sowing, and spring weeding of the fields. All of this caused a considerable rise in the productivity on the fields of the experiment stations in comparison to that of the peasants and even of the estates. But it must still be recognized that fluctuations did take place, just the same, about this considerably higher level, under the influence of the same natural forces as on the prerevolutionary peasant farms and estates. However, the better cultural practice and the higher level of productivity led to important practical consequences: the same yield of spring wheat was obtained from the experiment station fields under considerably worse meteorological conditions as was obtained by the peasants and estates under considerably better conditions. This thesis has been unconditionally proved by me through various statistical groupings. The reduction

in relative variability (such as it was) must likewise be attributed to the achievements of the experiment stations of the arid Povolzh'ye.

One of the major accomplishments of the Saratov experiment station was the application of snow-retention, as a special method of agricultural engineering for increasing soil moisture stocks by the beginning of the spring-summer period. The eighteen-year experience of the Saratov station showed that snow-retention considerably augments the success of the struggle against drought. It acts strongly to increase the yield, which reaches 14.61 tsentner per hectare as against 10.91 tsentner without snow-retention, and reduces the dependence of the yield on the spring-summer weather conditions; with snow-retention, the correlation coefficient between yield and rainfall from germination to the end of ripening is ' 0.729, while for the same eighteen years the closeness of the same connection, without snow-retention, is measured by the correlation coefficient ' 0.818.

In the years of worst drought, the plots with snow-retention gave relatively excellent yields for those years. Even in the catastrophic year of absolute drought, 1921, when absolute drought conditions prevailed at Saratov station throughout the entire life of spring wheat, from germination to maturation, the use of snow-retention assured a crop of 5.70 tsentner per hectare (even somewhat higher than the mean prerevolutionary peasant yield) while the spring wheat on the other plots of the experimental farm perished. In the arid years 1924 and 1927, the Saratov station had respective yields of 4.75 and 5.73 tsentner per hectare without snow-retention, against 10.0 and 8.0 tsentner per hectare with snow-retention.

The effectiveness of snow-retention is apparent from the following table, giving comparison of the spring wheat yield with and without snow-retention in groups of years with varying rainfall.

[see Table 10 on following page]

It is evident from this table that yield increases with rainfall both on plots with and plots without snow-retention, but that the yield in all groups from plots with snow-retention exceeds that from plots without it.

Thus the use of snow-retention in the arid zone makes it possible to increase the average (and precisely the average) productivity of spring wheat to fifteen tsentner per hectare (even without including that crop in a fertilized crop rotation or without special fertilizing of spring wheat) and weakens the dependence of yield on the spring meteorological factors. This great significance of snow-retention was taken into account by our socialist fatherland and is widely used, on the initiative of the party, on kolkhoz and sovkhoz fields. The relative variability of yield is also reduced in comparison to the Saratov station yields without snow-retention.

Thus, for instance, the productivity of the sixth group is 4.1 times that of the first group in the experiments without snow-retention, and only 2.1 times that of the first group with snow-retention. When three groups of six years each are used instead, the respective indices are 2.4 and 1.84. Even a certain reduction of absolute variability is apparently observed with snow-retention. The range between the first and sixth groups, when six groups of

TABLE 10

YEARS GROUPED BY RAINFALL FROM
GERMINATION TO END OF RIPENING

YIELD OF SPRING WHEAT AT SARATOV EXPERIMENT STATION
IN TSEENTNER PER HECTARE

	WITHOUT SNOW RETENTION		WITH SNOW RETENTION	
3 years with lowest rainfall (0 - 45 mm.)	3.72		8.55	9.71
3 years with rainfall 46 - 65 mm.	7.93	5.82	10.88	
3 years with rainfall 66 - 91 mm.	10.41		14.01	15.75
3 years with rainfall 92 - 109 mm.	13.07	11.74	17.49	
3 years with rainfall 110 - 120 mm.	12.67		17.10	17.85
3 years with highest rainfall (over 120 mm.)	15.20	13.93	18.60	

three years each are used, is 11.50 tsentner without snow retention and 10.05 tsentner with it. With three groups of six years each, the absolute ranges are equal: 8.15 and 8.14. (For all the variants discussed up to now, the absolute variability increased somewhat with rising level of cultural practice. All that has been set forth above allows us to define the eighteen-year experience of the Saratov Station with snow retention (to supplement the other methods of spring wheat culture at that station) as a fourth level of cultural practice in the arid Povolzh'ye. Snow retention, however, only weakens the dependence of yield on weather, without liquidating it. The use of artificial irrigation, as is stated in the above-cited work of Shneydyuller, which is based on the materials of the twenty-three-year experiments of the Trans-Volga Kostychev experiment station, even under the very imperfect methods of irrigation used at that station, in the first place immensely increases the mean yield (the long term irrigation was 18.9 tsentner, while the control plots without irrigation yielded only 5.70 tsentner in the second place, it almost entirely liquidates the dependence of yield on weather. The correlation coefficient between the yield of irrigated spring wheat and the May and June rainfall is almost zero (+0.03), while the same coefficient on unirrigated plots reaches the value of +0.644. The harmful action of high temperatures and dry winds is also considerably weakened by irrigation.

[see Table 11 on following page]

By increasing the atmospheric humidity and reducing the temperature, irrigation exerts a favorable influence on the micro-climate of the sown tracts.

TABLE 11

PLOTS	CORRELATION COEFFICIENT OF YIELD	
	WITH THE NUMBER OF DAYS WITH HARMFUL TEMPERATURE, 30 DEGREES AND OVER, DURING MAY AND JUNE	WITH THE DRY- WIND INDEX
With irrigation	-0.400	-0.184
Without irrigation	-0.673	-0.439

Experiments conducted in 1934 by the All-Union Institute of Grain Farming in the Trans-Volga, on the Yershov tract, showed that when the irrigation system was rationally organized, yields considerably in excess of those from the Kostychev experiment station could be obtained.

Twenty-four tsentner per hectare were obtained with two irrigations, without fertilizer. The control plot gave 10.5 tsentner (approximately as much as the Saratov experiment station obtained from plots without snow retention).

Thus we have the fifth level of cultural practices: irrigated farming without fertilizer.

The use of fertilizer in irrigated farming raises the level of cultural practice still higher, to the sixth level. The results of the use of fertilizer in irrigated farming are as follows.

The experiment in 1934 by the All-Union Institute of Grain Farming at Yershov, using two irrigations, yielded 37.1 tsentner per hectare.

The Stakhanovite V.I. Kursakov obtained 43.7 tsentner (See "Agricultural Technology of High Yields", published by the All-Union Academy of Agricultural Sciences in honor of Lenin, pages 34 and 55.) In the very arid year 1936, in the Trans-Volga, in the Saratov oblast, using four snow retentions and two irrigations.

Finally, the experience of the agricultural Stakhanovite V. G. Olesnikov stands entirely apart. He used no fertilizer, but carried out his sowings on plowed-up virgin soil. In the same year 1936, with careful cultivation of the land, weeding the sown

area until all weeds were completely wiped out, using two irrigations, he obtained a yield of forty-eight to fifty tsentner per hectare, (a method successfully applied by the Stakhanovite Yefremov in Western Siberia).

The system of irrigated farming thus completely revolutionizes the conditions of agriculture in the arid zone, carrying the yield up to levels unseen under the conditions of dry agriculture, while this yield, liberated from the caprices of the weather, acquires a stability which is important for the national economy. The results of the system of irrigated farming are a very high and stable yield.

On the fields of our socialist fatherland this system is already beginning to become reality. On the initiative of the party, an extensive plan has been drawn up for the irrigation of the Trans-Volga, the Marych steppe, North Caucasus, Transcaucasia, Central Asia, etc. -- and this plan is already being executed.

We shall now briefly indicate what seed material was used in all the cases we have considered. The estates used a better seed material than did the peasants. This superiority hardly went as far as to make the use of improved seed varieties general on the estates; but the considerably more widespread distribution there of winnows, sorting machines and grain cleaning machines led to their seeding cleaner, larger and heavier material than the peasants. The Bezenchuk station used the "Poltavka" variety. The work of Shneydyuller on the Saratov station covered the yields from sixty-two selected lines, from the prize variety-tests conducted by the All-Union Grain Growing Institute from 1915 to 1931. Thus when we consider the Saratov station's work, we are dealing

with seed material of improved varieties. The Melanopus 069 variety was seeded in the irrigation experiment of the Yershov experimental tract. V.I. Kursakov used Lutescens 062, and the Stakhanovite Olaynikov seeded Melanopus 069. Thus we have everywhere to do with experiments in which improved material was seeded.

Table 12 groups and arranges the fundamental conclusions of the present work.

[See Table 12 on following page]

But the norms of productivity for irrigated farming given in this table cannot be considered upper limits. Neither the 1934 experiments of the All-Union Grain Growing Institute, nor Kursakov, nor Olaynikov in his 1936 experiments, made use of certain agrotechnical methods that have been successfully employed by Stakhanovites of other oblasts and zones. Cross-sowing, which assures the uniform distribution of the seed throughout the field and assigns the same area of supply to each plant; increase of the rate of seeding or crop density (a method successfully applied by the Stakhanovite Yefremov in Western Siberia); the vernalization of seed -- all these are methods that should be very effective under the conditions of irrigated farming with fertilizer. But there are two dangers that must be noted with respect to irrigated farming -- dangers that not infrequently imperil the yield in the dry farming of the arid Povolzh'ye: that of lodging of the crop, and that of crop damage by certain pests (the Swedish fly, for example) and especially by parasitic diseases (notably rust). In its 1934 experiments the All-Union Institute of Grain Growing notes both of these dangers. May was hot in 1934; June was cool, with plenty of rainfall. In July, after heavy rains with high winds,

LEVELS OF CULTURAL PRACTICE		YIELD IN TSENTNER PER HECTARE	EXPLANATION
I LOWEST LEVEL OF CULTURAL PRACTICE	PREREVOLUTIONARY PEASANT FARMS		Immense influence of spring- summer weather on fluctuations in the yield of spring wheat (amount of rainfall, number of rainy days, temperature condi- tions, saturation deficit). But these fluctuations at dif- ferent levels of cultural prac- tice took place about levels of productivity which were likewise different, so that the practical importance of cultural practice was expressed in the same yield (and often even a larger one) being obtained by farms at higher technical levels, under considerably worse meteorolo- gical conditions, those ob- tained by farms at lower technical levels under con- siderably better meteorolo- gical conditions (cf. the se- ries of tables, especially Table 7. With the rise in cultural practice, the abso- lute variability or range in- creased, but the relative va- riability or coefficient of variation declined. This si- tuation is marked only in the comparison of Levels I and III of cultural practice.
	Former Saratov Guberniya	4.74	
	Former Samara Guberniya	5.44	
II LEVEL OF CULTURAL PRACTICE	ESTATES		
	Former Saratov Guberniya	6.84	
	Former Samara Guberniya	6.95	
III LEVEL OF CULTURAL PRACTICE	25-year experience of Bezenchuk Station	9.25	
	18-year experience of Saratov Station	10.91	
IV LEVEL OF CULTURAL PRACTICE	18-year experience of Saratov station with snow retention	14.61	IV Level. Influence of spring- summer weather is weakened. Relative variation is strongly reduced, and a tendency towards reduction of absolute varia- bility is also noted.

[Table 12 continued
on following page]

TABLE 12 (continued)

V LEVEL OF CULTURAL PRACTICE	<u>Irrigated farming without fertilizer</u> 23-year experience of Kostychevsk Station	18.90	The influence of rainfall on the fluctuation of the yield is annulled. The influence of high temperatures and dry winds is strongly attenuated. It must be assumed that as the system of irrigation is im- proved (VIZKh experiments) and the use of fertilizer in- troduced in irrigated farming (Level VI) even these remnants of the influence of spring- summer weather (soil and at- mospheric drought) will be wiped out.
	Experience of VIZKh at Yershov, 1934	24.00	
VI LEVEL OF CULTURAL PRACTICE	<u>Irrigated farming with fertilizer</u> 1934 experience of VIZKh	37.1	
	Experience of V.I. Kursakov	45.7	
	Experience of V.G. Oleynikov (It appears that Oleynikov used no fertilizer, but sowing on newly plowed virgin land is equiva- lent to sowing on fer- tilized soft earth)	48-50	

which occurred during the time of full ripening, the wheat crop partially lodged down, especially on the tracts that had had three irrigations (emphasis ours). "Stem rust mostly affected the irrigated tracts, especially those irrigated three times" (emphasis ours). There is thus nothing surprising in the fact that thrice-irrigated and fertilized tracts not only failed to give any additional yield but even displayed a tendency towards reduced yield; in the 1934 VIZKh experiments, plots with three irrigations yielded 36.3 tsentner per hectare, and those with two irrigations 37.1 tsentner. In the dry year 1936 the Stakhanovite V.I. Kursakov, on the suggestion of the Saratov Institute of Grain Growing, ran a three-irrigation patch on a plot beside a two-irrigation one. He obtained good results: the extra irrigation gave four to five more tsentner per hectare.

The problem of seed selection in the fight against lodging and susceptibility of spring wheat to parasitic diseases is to develop non-lodging varieties, immune to disease. This task must be solved without delay, since the time is not far distant when the realization of the "Great Volga" projects will irrigate millions of hectares in the Trans-Volga, thus producing a great demand for non-lodging and immune varieties of spring wheat.

But great meliorative works must precede the application of irrigation. Even when the Great Volga project becomes reality, the entire acreage under cultivation cannot be irrigated. Therefore, the method of using snow retention to increase soil moisture stocks takes on particular interest. This method of agricultural engineering, in essence, already constitutes a beginning of irrigation (and a very effective one at that). It requires no meliora-

tive works and is accessible to every kolkhoz. It has been tried out over the eighteen-year experience of the Saratov agricultural station. That period included 1921, a year of absolute drought, 1924 and 1927, very dry years, as well as years of average and abundant rainfall. As we have seen, snow retention, without the use of fertilizer, increased the yield to 14.61 tsentner per hectare. Further increase of productivity on the snow-retention plots demands, first of all, the use of fertilizer on those plots. (We shall not touch the question what method of fertilizer application would be most effective and most productive with respect to those stocks of fertilizers of which the kolkhozes dispose at the present time (manure, ashes, nightsoil and the relatively small amount of mineral fertilizer which can be allocated to the grain growers at the present moment). Three methods of applying fertilizer to spring wheat may be noted:

1. Inclusion of spring wheat in a fertilized crop rotation;
2. Introduction of fertilizer in autumn during the shallow winter breaking [? zyblovoy vspashka]
3. Supplemental spring fertilizing of the spring sowings.

This method has been very effectively employed by many Stakhanovites of agriculture. Apparently the most productive method would be inclusion of the wheat in a fertilized crop rotation, plus supplementary fertilizing.

We have seen that the use of fertilizer in irrigated farming increased the yield by fifty-five percent (from 24 to 37.1 tsentner). Unfortunately the eighteen-year test of snow retention without fertilizer at the Saratov station was not accompanied by a parallel test of snow retention with fertilizer. (Apparently parallel experiments were made by the Saratov experiment station since 1936

on fertilizer with snow retention, but I do not dispose of the results.) However, if an increase in productivity be assumed, equal to half the effect given by fertilizer in irrigated farming, i.e., if we assume that fertilizer would increase the productivity by twenty-seven percent, then the presumptive average productivity (I must emphasize the word average) is increased to at least 18.5 tsentner per hectare. (In 1936 the Stakhanovite Chumakov, using snow retention in Western Siberia, manure fertilizer, sowing in widely spaced rows and giving two cultivations, obtained a yield of 30 tsentner (Teesium variety) per hectare from soft earth. Against this background the use of supplemental fertilizer (night-soil) increased the yield to 50 tsentner. ("Agricultural Technology of High Yields", page 333). P.F. Yemel'yanov, chief agronomist of the Western Siberia Oblast, considers snow retention to be of exceptional importance for a considerable portion of the kray (op. cit., page 334). He also notes the same exceptional importance of snow retention for winter wheat as well (page 335)). (If fertilizer is equally effective on irrigated plots and snow-retention plots, we would have a yield of 22.5 tsentner on the latter). Use of cross-sowing, increasing the crop density, introduction of correct crop rotations, including sown grasses, opens up broad perspectives for the further enhancement of productivity on snow-retention plots.

SUPPLEMENT

Up to now we have been speaking of the yield of spring wheat in dry farming, with snow retention, or on irrigated tracts. In his above-cited work, V. Shneydyuller investigates the influence of the spring meteorological factors on the yield of winter wheat in dry farming at the Saratov station in irrigated farming at the Kostychev experiment station. We present the results of his work, with my own supplements thereto, on the yield of winter wheat from snow-retention plots.

The work cited states that the correlation coefficient between the May-June rainfall and the yield for winter wheat is $+0.826$, and that between harmful temperatures and the same yield it is -0.840 . The joint influence of these two factors is expressed by a correlation coefficient equal to 0.92 . The mean productivity of winter wheat at the Saratov station, however, was fairly high: it was 17.7 tsentner per hectare, ranging from 7.6 to 28.6 tsentner (except for 1921 when the crop failed). Just as for spring wheat, snow retention, which increases the soil moisture stocks by spring, and, for good measure, also assures an optimum snow cover for winter wheat, proved to be a very effective agricultural measure, increasing the productivity of winter wheat to 22.2 tsentner per hectare (i.e., by 4.5 tsentner), with a range from 10.5 to 32.8 tsentner per hectare. But snow retention, while it increases the productivity in the arid zone, does not completely solve the question of liquidating the drought. Just as for spring wheat, it is irrigation that introduces this revolution into the agriculture of the arid zone.

In spite of the fact that the meteorological and soil conditions are worse at the Kostychev experiment station than at Saratov, and in spite of the imperfect system of irrigation, the mean twelve-year productivity of winter wheat from the irrigated plots at Kostychev was 25.6 tsentner per hectare. It would, however, be more correct to compare the yield for years in which there is data both for irrigated plots (Kostychev station) and for snow-retention plots (Saratov station). The seven-year period 1924 - 1930 gives a yield of 26.6 tsentner for winter wheat in irrigated farming at Kostychev, and 19.6 tsentner for winter wheat on the snow-retention plots of the Saratov station (a difference of seven tsentner in favor of irrigated farming).

Further, as the Shneydyuller study shows, the yield on the irrigated tracts of Kostychev: (1) are completely emancipated from the influence of the rainfall (the correlation coefficient with the May-June rainfall is +0.03 there, while it is +0.826 at the Saratov station); are subject to a considerably lesser extent to the harmful influence of high temperatures (-0.16 at Kostychev, under irrigation; -0.78 at Saratov station). In a word, the picture is the same for both spring and winter wheat.

APPENDIX 1 to CHAPTER II

SUMMARY TABLES OF THE METEOROLOGICAL INDICES, IN THE SEVERAL VEGETATIVE PERIODS, OF SPRING
WHEAT, ON THE PEASANT FARMS IN THE SOUTHERN PORTION OF FORMER SAMARA GUBERNIYA

Year	1. MEAN NUMBER OF RAINY DAYS (d)									Yield Expressed as Deviation from the Trend (y)
	1	2	3	4	5	6	7	8	9	
	d ₁	d ₂	d ₃	d ₄	d ₅	d ₆	d ₇	d ₈	d ₉	
1886	1.1	3.6	1.7	1.7	<u>7.6</u> 3.8	3.3	0.8	5.9	4.0	+12.2
1887	2.1	0.4	1.1	1.9	<u>1.2</u> 1.2	1.9	2.8	3.3	2.2	- 6.9
1888	0.6	0.5	0.2	2.1	<u>5.0</u> 2.5	1.1	2.0	0.6	5.1	- 5.7
1889	1.1	0.4	0.4	0.8	<u>3.6</u> 1.8	3.1	1.9	1.7	0.4	- 5.9
1890	1.7	0.8	0.2	1.5	<u>0.9</u> 0.5	0.4	4.3	3.8	2.0	- 6.6
1891	1.6	1.5	0.7	2.0	<u>2.2</u> 1.1	0.9	0.9	0.6	1.3	-15.3
1892	0.3	0.6	0.4	4.0	<u>4.8</u> 2.4	2.1	2.7	2.0	0.2	+10.3
1893	2.7	0.5	3.1	0.0	<u>0.8</u> 0.8	2.9	1.7	2.4	1.4	+ 2.7
1894	3.6	0.0	2.3	3.1	<u>4.2</u> 4.9	4.4	3.0	1.0	2.2	+25.0
1895	0.5	0.6	1.1	3.7	<u>5.6</u> 2.8	1.0	3.3	2.1	1.4	- 2.5
1896	1.6	2.1	3.8	1.1	<u>4.7</u> 4.7	4.0	4.2	3.8	2.3	+19.3
1897	1.4	3.4	1.1	1.0	<u>2.5</u> 1.3	2.1	2.0	2.7	0.3	- 3.4
1898	3.2	0.2	0.7	1.7	<u>0.3</u> 0.3	2.4	1.9	0.6	4.5	-10.3

1. MEAN NUMBER OF RAINY DAYS (d) (continued)

Year	1	2	3	4	5	6	7	8	9	Yield
1899	2.1	1.8	1.8	1.9	<u>0.9</u> 0.9	0.2	2.8	2.0	0.4	- 1.9
1900	0.9	1.3	2.4	1.3	<u>1.7</u> 1.7	5.6	1.2	4.3	2.3	+ 7.6
1901	1.7	1.0	1.2	2.0	<u>1.5</u> 1.5	1.8	0.6	0.1	3.0	-11.1
1902	2.6	2.4	2.0	0.8	<u>3.1</u> 3.1	1.7	1.8	1.2	1.6	+ 3.9
1903	0.2	0.0	2.1	5.7	<u>2.9</u> 2.9	0.6	0.5	0.2	1.1	- 1.3
1904	0.1	0.2	1.1	4.4	<u>1.5</u> 8.6	2.4	1.5	1.6	1.1	+17.1
1905	1.3	1.0	0.4	0.3	<u>4.3</u> 2.3	1.2	1.5	3.9	2.2	- 5.7
1906	1.3	0.5	0.3	0.5	<u>0.7</u> 0.7	0.8	1.8	5.1	1.1	-20.5
1907	0.4	0.8	3.1	2.3	<u>0.7</u> 0.7	2.2	1.0	1.5	2.2	- 9.9
1908	3.3	3.3	3.0	1.3	<u>1.0</u> 1.0	1.2	1.4	0.7	0.7	- 9.8
1909	1.6	1.7	2.0	3.9	<u>4.4</u> 2.2	3.8	0.9	1.4	1.5	+11.8
1910	2.8	2.6	1.6	3.4	<u>0.8</u> 0.8	1.1	2.9	1.1	0.9	+ 3.6
1911	0.0	0.5	0.7	0.9	<u>4.4</u> 4.4	2.2	0.0	0.4	2.7	-22.0
1912	0.9	1.7	2.8	2.6	<u>0.6</u> 0.6	1.3	1.3	4.0	0.8	- 8.8
1913	2.3	0.9	2.5	2.1	<u>2.4</u> 3.1	1.9	1.7	3.8	5.0	+15.4
1914	3.0	1.9	0.5	1.0	<u>7.1</u> 3.6	0.5	1.5	0.3	1.2	+ 2.3
1915	1.7	3.2	3.8	3.6	<u>4.6</u> 2.3	2.8	0.7	1.7	0.8	+16.4

2. RAINFALL (p)

VEGETATIVE PERIOD

Year	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉
1886	5.3	30.0	7.1	10.1	<u>63.5</u> 31.8	24.9	6.2	57.2	35.0
1887	7.6	3.0	18.4	8.9	<u>5.3</u> 5.3	7.1	18.9	29.4	13.4
1888	1.5	1.9	1.2	10.2	<u>31.3</u> 15.7	4.6	13.0	3.2	40.2
1889	3.7	3.8	3.1	2.8	<u>21.9</u> 11.0	13.7	7.9	7.2	1.1
1890	4.2	4.8	0.2	5.4	<u>5.0</u> 2.5	4.6	35.4	26.2	12.7
1891	8.3	5.7	6.3	11.8	<u>7.9</u> 4.0	4.1	5.2	2.4	11.5
1892	0.8	2.1	1.1	31.8	<u>25.5</u> 12.8	8.3	21.3	19.3	0.2
1893	7.9	2.0	14.4	0.1	<u>8.6</u> 8.6	13.6	16.3	28.0	13.5
1894	14.0	0.3	11.8	15.5	<u>21.3</u> 21.3	22.8	9.8	4.5	12.5
1895	2.4	2.3	5.6	19.1	<u>30.6</u> 15.3	3.5	18.2	13.8	12.8
1896	7.9	18.4	38.0	4.5	<u>36.5</u> 36.5	20.7	27.1	21.0	22.4
1897	3.2	25.7	10.3	2.3	<u>15.5</u> 7.8	11.7	6.9	15.9	1.0
1898	15.0	0.7	1.6	8.0	<u>0.3</u> 0.3	16.7	10.2	3.5	8.3
1899	12.6	13.2	6.6	10.8	<u>2.4</u> 2.4	1.1	12.2	22.8	2.1

2. RAINFALL (p) (continued)

1900	2.4	9.0	15.7	5.8	14.7	32.9	10.1	26.1	8.1
1901	12.6	6.8	4.5	14.8	14.7	9.2	1.8	0.6	9.9
1902	16.1	13.4	18.7	3.4	8.5	17.0	5.9	4.1	11.3
1903	0.4	0.1	14.5	36.3	23.4	3.0	3.7	0.4	6.2
1904	0.3	0.8	4.9	25.9	23.4	18.1	12.2	10.1	6.2
1905	5.0	6.8	1.4	1.2	25.8	10.3	5.9	20.4	9.5
1906	4.1	1.7	0.5	4.3	12.9	3.8	12.8	39.6	13.9
1907	1.5	2.4	15.2	7.9	49.6	28.9	9.7	9.9	24.7
1908	20.5	16.9	11.3	5.3	24.8	4.4	5.7	3.7	1.9
1909	6.3	5.2	12.7	16.3	10.7	24.8	7.1	4.1	5.4
1910	13.6	15.3	14.5	13.5	2.2	6.1	22.0	6.0	3.2
1911	0.2	1.8	2.3	3.2	1.7	20.6	0.0	1.1	11.0
1912	3.4	15.3	20.5	21.6	5.4	7.4	13.0	43.7	5.9
1913	13.0	2.5	14.5	23.4	24.8	9.2	10.0	26.2	38.9
1914	14.6	6.9	3.4	11.9	12.4	3.9	6.2	1.1	5.3
1915	6.9	33.5	21.1	24.2	10.0	15.7	4.6	30.9	2.4

3. SATURATION DEFICIT ($E - e$)

VEGETATIVE PERIOD

Year	($E - e$) ₁	($E - e$) ₂	($E - e$) ₃	($E - e$) ₄	($E - e$) ₅	($E - e$) ₆	($E - e$) ₇	($E - e$) ₈	($E - e$) ₉
1886	8.4	9.1	10.8	14.9	11.2	8.4	15.5	7.8	9.7
1887	7.4	11.9	18.4	8.9	15.3	11.9	13.2	8.5	11.3
1888	7.2	10.6	18.0	11.2	10.4	12.8	11.5	24.5	10.2
1889	6.7	13.7	11.7	16.6	11.9	15.4	14.5	19.5	22.4
1890	3.9	7.1	11.7	10.3	12.2	25.3	12.3	13.1	18.4
1891	5.4	8.1	7.7	9.1	17.6	23.6	19.5	28.8	23.5
1892	7.0	14.1	10.7	10.3	16.7	15.0	12.7	20.1	28.5
1893	3.8	7.8	7.3	13.3	21.4	11.3	15.6	12.4	23.4
1894	8.0	14.0	13.9	9.2	8.7	10.2	8.4	16.1	15.7
1895	6.2	9.0	11.1	7.9	13.8	19.3	14.6	16.0	18.0
1896	8.7	7.9	9.7	11.5	8.1	9.9	10.2	10.6	10.0
1897	4.6	5.1	9.3	13.1	15.9	14.2	16.7	11.3	23.5
1898	3.5	8.1	19.1	14.7	13.1	12.7	15.3	24.4	19.5
1899	6.1	6.9	9.7	10.5	10.8	25.5	13.3	11.8	16.9
1900	9.0	9.0	11.0	10.9	15.3	7.3	16.6	17.6	11.6

3. SATURATION DEFICIT (E - e) (continued)

1901	5.6	10.8	7.7	9.8	13.5	11.2	18.4	21.5	16.8
1902	3.4	5.6	7.6	13.1	11.0	21.7	12.5	13.5	14.0
1903	5.5	10.1	8.9	4.9	10.2	12.9	15.1	21.1	21.6
1904	11.6	10.8	15.4	4.4	7.2	10.4	15.8	12.0	14.9
1905	11.2	8.4	11.3	16.0	12.3	13.3	17.6	11.6	10.5
1906	6.7	12.8	14.7	14.9	19.6	19.5	15.4	9.3	24.4
1907	8.0	9.8	8.5	7.7	12.5	13.1	17.0	25.5	16.1
1908	4.5	7.1	7.4	13.0	15.8	22.2	15.2	18.5	18.7
1909	7.5	13.5	13.2	5.9	14.9	13.1	15.5	19.4	15.6
1910	12.4	11.7	11.8	7.2	12.5	15.1	9.7	21.0	17.5
1911	8.5	9.8	10.7	16.7	9.2	13.7	29.0	35.3	13.4
1912	7.5	5.4	7.6	10.3	14.7	23.2	16.7	8.7	12.4
1913	6.4	8.9	5.2	5.7	9.6	10.4	14.9	13.7	8.6
1914	2.7	5.6	10.8	18.0	8.3	12.8	16.1	19.9	16.0
1915	9.1	6.4	4.9	6.4	10.0	10.8	13.2	13.4	19.7

Footnote to tables 1, 2 and 3

Two figures are given for the fifth period in Tables 1 and 2. The numerator is the total number of rainy days or amount of rainfall, as the case may be, for the decades composing the period, while the denominator is the mean number of rainy days or mean amount of rainfall per decade. In Table 3 the mean indices of saturation deficit for one decade are given for the fifth period.

Editor's Note

The content and method of construction of column d_5 in Table 1 and p_5 in Table 2 require certain explanations, which we have prepared on the basis of the unpublished work of V.M. Obukhov entitled "The Dependence of Yield on Meteorological Factors in the Trans-Volga of the Middle and Lower Volga from 1866 to 1915".

The author set himself the task of approaching as closely as possible to the basic vegetative periods of each year, using the data of the zemstvo statistics and the former Department of Agriculture ("...Year in Agriculture"), from which information on the beginning of sowing and harvesting of the spring cereals for each year could be extracted. If the decades of the growing season are counted forward from the beginning of sowing, or backwards from the harvest, then, in consequence of the varying duration of the growing season (from 9 to 11 decades), the decade with the same serial number will fall in different phases of plant development in different years. For this reason the author had recourse to counting off the decades from both ends, for four decades each, so that the central interval, designated d_5 or p_5 , consisted in some years of a single decade, in other

years, of two decades, and in certain years even of three, depending on the total duration of the growing season in that particular year. To make the characteristics of this central interval comparable from year to year, the number of rainy days (d) or the amount of rainfall (p) (the numbers above the line in the tables) were reduced to their average ten-day values, which are shown beneath the line.

Thus there were nine periods marked out, each of which lasted a decade, except for the fifth.

- | | |
|-------------|---|
| 1st period: | decade of sowing |
| 2nd period: | germination |
| 3rd period: | tillering, onset of shooting |
| 4th period: | |
| 5th period: | before heading (the end of the period is the onset of heading) |
| 6th period: | heading (sometimes the onset of blooming occurs in the second half) |
| 7th period: | blooming (sometimes the end of heading in the first half) |
| 8th period: | ripening |
| 9th period: | maturation |

APPENDIX 2 to CHAPTER II

CALCULATION OF THE SPRING WHEAT YIELD FROM PEASANT FARMS IN THE SOUTHERN PART OF THE FORMER SAMARA GUBERNIYA BY THE REGRESSION EQUATION, AND COMPARISON WITH THE ACTUAL YIELD

y : actual yield expressed as deviation from the trend line;

y_d : indices of the deviation of the yields calculated by the equation: $y_d = 35.0369 + 2.5969$, where $1D_7$ is the number of rainy days from the beginning of sowing to blooming inclusive;

$1y_E$: indices of yield, computed by the equation:

$$1y_E = -0.848610 - 0.150422 1H_7 + 0.005901 1H_7^2 - 0.004389 1H_7^3;$$

$$2y_E = 3.238396 - 0.316290 H_3 - 0.0077328 H_3^2;$$

y' : indices of deviations of the yields determined by the regression equation in which the indices y_d , $1y_E$ and $2y_E$ enter,

$$y' = 0.69541 y_d + 0.35441 1y_E + 0.283274 2y_E; R = 0.955;$$

$y - y'$: divergences between the indices of yield calculated according to the foregoing resultant equation and the actual deviations of the yield from the trend line, i.e., the difference $= y - y'$.

Year	y	y_d	$1y_E$	$2y_E$	y^1	$y - y^1$
1886	+12.2	+15.4	+ 0.1	- 0.3	+10.6	+ 1.6
1887	- 6.9	- 6.0	+ 0.1	+ 0.4	- 4.0	- 2.9
1888	- 5.7	- 5.7	+ 3.0	- 3.6	- 3.9	- 1.8
1889	- 5.9	- 6.2	- 2.0	+ 1.9	- 4.5	- 1.4
1890	- 6.6	-10.1	- 3.1	+ 3.3	- 7.2	+ 0.6
1891	-15.3	-10.1	-19.7	-11.5	-17.3	+ 2.0
1892	+10.3	+ 2.9	- 1.0	+ 1.4	+ 2.0	+ 8.3
1893	+ 2.7	- 5.2	- 3.9	+ 3.1	- 4.1	+ 6.8
1894	+25.0	+20.7	+25.0	+ 3.4	+24.2	+ 0.8
1895	- 2.5	+ 5.2	- 1.2	+ 3.5	+ 4.2	- 6.7
1896	+19.3	+19.7	+14.5	+ 2.2	+19.5	- 0.2
1897	- 3.4	- 0.6	- 2.5	+ 2.6	- 0.6	- 2.8
1898	- 10.3	- 8.5	- 1.2	- 3.5	- 7.3	- 3.0
1899	- 1.9	- 5.7	- 2.5	+ 2.8	- 4.1	+ 2.2
1900	+ 7.6	+ 1.6	+ 0.1	+ 3.0	+ 2.0	+ 5.6

(continued)

1901	-11.1	-10.1	- 0.7	+ 0.2	- 7.2	- 3.9
1902	+ 3.9	+ 1.6	- 2.0	+ 3.4	+ 1.4	+ 2.5
1903	- 1.3	- 4.5	+ 7.4	+ 0.6	- 0.3	- 1.0
1904	+17.1	+11.6	+21.1	+ 2.9	+16.4	+ 0.7
1905	- 5.7	-14.7	- 2.5	+ 2.7	-10.3	+ 4.6
1906	-20.5	-20.0	-19.7	+ 1.2	-20.6	+ 0.1
1907	- 9.9	- 8.3	- 0.2	- 5.2	- 7.3	- 2.6
1908	- 9.8	+ 1.9	-11.4	+ 2.5	- 2.0	- 7.8
1909	+11.8	+11.6	+ 0.1	+ 2.0	+ 8.7	+ 3.1
1910	+ 3.6	+ 3.7	+ 1.2	+ 0.7	+ 4.2	- 0.6
1911	-22.0	-12.9	-16.6	-28.8	-23.0	+ 1.0
1912	- 8.8	- 6.5	- 7.6	+ 0.6	- 7.0	- 1.8
1913	+15.4	+17.9	+11.7	+ 3.5	+17.6	- 2.2
1914	+ 2.3	+ 4.4	- 1.2	+ 1.6	+ 3.1	- 0.8
1915	+16.4	+10.9	+11.7	+ 3.4	+16.8	- 0.4

(From a manuscript completed in 1940)

Introduction.

217

stocks in the spring; and (d) under irrigation, which, by extreme increases in the yield, completely liberates it from the influence of rainfall, and very powerfully softens the action of high temperatures, saturation deficit and dry winds. It was possible to trade the influence of insect pests and fungous diseases on the yield of wheat, as well as the wintering conditions in the case of winter wheat, only for the Poltava station. The reason was the same: the necessary materials were not received. The author of the present work has supplemented his own investigations by references to published and unpublished works of other authors. (These works are indicated at the corresponding places in this text.

As a result the long-term data from the following stations have been used:

For spring wheat: Kupinsk experimental field (West Siberia), Omsk, Chishma, Bezenchuck, Saratov, Kostychev experiment station, Balashov, Sinel'nikovo, Poltava, Sumy.

For winter wheat: Rostov-on-Don, Yeysk, Krasnodar, Sinel'nikovo, Odessa, Poltava, Sumy, Shatilovskaya experiment station, Saratov, Kostychev experiment station.

I. SPRING WHEAT

The arid steppe portion of the USSR was studied from the materials of five agricultural stations: Saratov, Kostychev, Bezenchuck, Sinel'nikovo and Kupinsk.

The works at all of these stations speak of the immense importance of the spring-summer weather for the yield of spring

wheat in the arid zone. Rainy, cool weather with a low saturation deficit results in upward deviations of the yield; rainless weather, high temperatures, low atmospheric humidity lead on the contrary to reduced yields. The correlation coefficients of the spring wheat yield with the spring-summer rainfall at the above stations are as follows:

TABLE 1

CORRELATION COEFFICIENTS BETWEEN YIELD OF SPRING
WHEAT AND SPRING-SUMMER RAINFALL

Saratov	+ 0.775 (tillering to heading)
Bereznyak	+ 0.925 (from 3rd April decade to 3rd June decade inclusive)
Sinel'nikovo	+ 0.504 (sowing to heading)
"	+ 0.747 (tillering to onset of blooming)
"	+ 0.498 (blooming to maturation)
Kupinsk	+ 0.900 (June and 1st July decade)

The degree of humidity of the weather is characterized not only by the amount of rainfall but also by high temperatures and saturation deficit: the higher the temperature, and the greater the saturation deficit, the more intensively do the processes of drying out the soil take place, both directly and through the intensified transpiration by plants, and the more intensive is the overheating of the plants. All of the investigators, therefore, have used an index of atmospheric drought in one form or another in addition to the rainfall.

At the Saratov experiment station this was done by calculating the number of humid days for each year, a day being considered humid if the evaporation power, computed by the formula $\frac{1}{2} (E - e)$, (E is the absolute atmospheric humidity that would saturate the space / at the given temperature /, e is the actual absolute atmospheric humidity, and $E - e$ is the saturation deficit.) did not exceed the amount of rain falling or remaining over from previous days. This index (the number of humid days) thus is a composite one taking into account both rainfall, temperature (since E is a function of the temperature), and the atmospheric humidity. The correlation coefficients between yield and the number of humid days during the several periods are as follows: tillering to heading, + 0.847; germination to tillering, + 0.594; heading to maturation, + 0.440. V. Shneydyuller, the author of the work on the Saratov experiment station, also included a dry-wind index, consisting of the accumulated wind velocities at 1300 hours of the dry days during the separate periods in different years. Inclusion of this factor brings the correlation coefficient for the Saratov experiment station up to 0.94.

Thus the amount of rainfall during the phase from tillering to heading, the saturation deficit for the same period, and the dry winds, are the principal factors in the fluctuation of the yields at Saratov experiment station.

At the Bezenchuk station (Krybyshev oblast) we included as a supplementary factor the intensity of evaporation as measured by Vil'd evaporimeter. Inclusion of this factor decreases the correlation coefficient for Bezenchuk station to 0.950 (The evaporation by the Vil'd evaporimeter, as follows, both from Dalton's law and experimental verification, depends primarily on the saturation deficit

(E -e). I have, moreover, established in one of my unpublished works "On the influence of meteorological factors on the variation of soil moisture stocks", that a shuttered instrument shelter in which an instrument is placed does not isolate it completely from the influence of the wind, and that the wind force is very appreciably reflected in the instrument readings. Thus our work at Bezenchuk took account of the influence of rainfall, saturation deficit and wind force. (cf. diagram 9 in the preceding article, p. 186).

At the Kupinsk station (on the West Siberian steppe), the author of the work, Comrade Dyakin, notes the negative influence of the saturation deficit (correlation coefficient 0.75). However the author has not set up the regression equation combining the action of both rainfall and saturation deficit; but in any case the correlation coefficient for the joint influence of these factors cannot be less than 0.90.

Finally, the Sinel'nikovo experiment station (Dnepropetrovsk oblast) gives a correlation coefficient of 0.917. This coefficient is obtained as a result of the combined action of three factors: the percentage of days with useful rainfall (1) from sowing to tillering; (2) from tillering to blooming and (3) from blooming to maturation, with a day counted as having useful rainfall if the amount of rainfall exceeded the evaporimeter reading (This index - the number of days with useful rain -- takes into account both the raininess of the weather and the saturation deficit and, particularly, the wind force (see the preceding footnote on the evaporimeter readings). for the same day (the details are set forth in the unpublished investigation of the author on the Sinel'nikovo station; see also Diagram 1).

Diagram 1

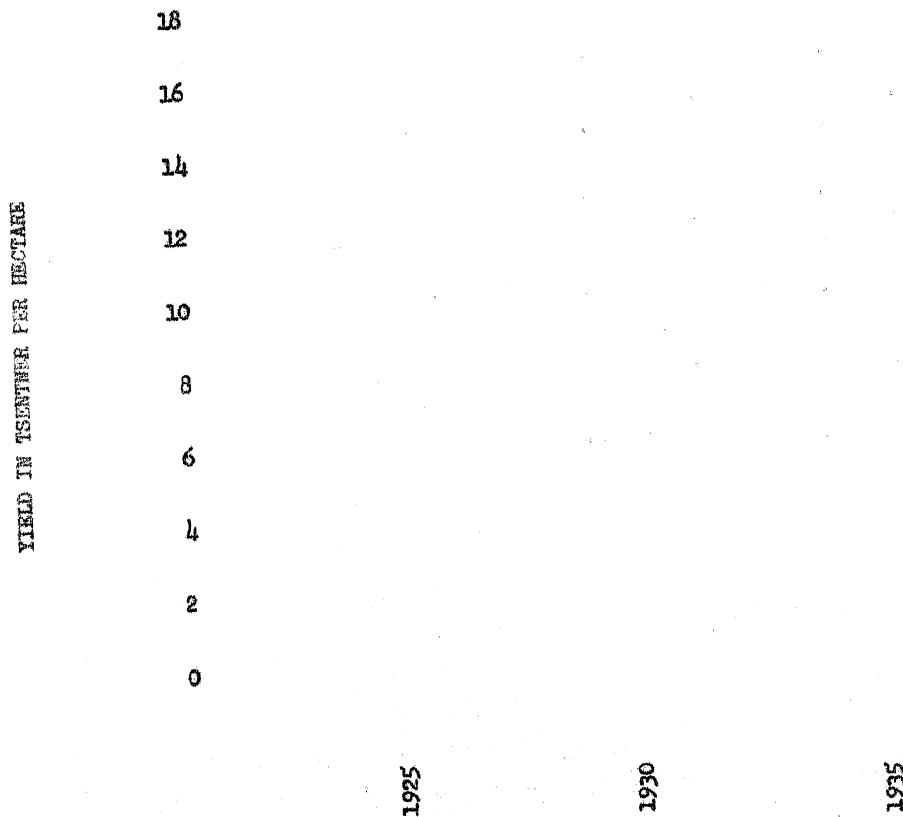
YIELD OF SPRING WHEAT AT SINEL'NIKOV AGRICULTURAL EXPERIMENT STATION

————— Observed yield

----- Yield calculated from regression equation:

$$i^y = 2.1930 + 0.1316 d_1 + 0.1862 d_2 + 0.2140 d_3;$$

$$R = 0.935 ; R_1 0.917$$



Note - The 1930 yield is not shown, since spring wheat in that year suffered badly from wheat rust.

Consequently the coefficients of correlation of the combined action of raininess and atmospheric drought at four stations in the arid zone have values exceeding 0.90, namely: Saratov, 0.910; Besenduk, 0.950; Kupinsk experiment station, not less than 0.9000, Sinel'nikovo, 0.917.

The absence of data from the Odessa experiment station, from which I received no materials, leaves a substantial gap in the southwestern steppe zone of the Ukraine. (I consider this gap substantial because the work of M. P. Kuritsyn "On the productivity of wheat", based on the spring wheat yield from peasant farms in the former Kherson guberniya in the prerevolutionary epoch, poses curious questions for the southwestern steppe of the Ukraine. His investigation notes the predominantly positive influence of the rains preceding the autumn harvest, the negative influence of the temperature during the same period, and the relatively weak effect of the spring rains.)

The correlation coefficient between yield and the rainfall from the 3rd October decade to the first December decade inclusive + 0.568

The correlation coefficient between yield and temperature from the 3rd October decade to the 1st December decade inclusive - 0.681

The correlation coefficient between yield and rainfall from the 2nd April decade to the 1st May decade inclusive + 0.373

This indicates that a cool and rainy autumn on the southwestern steppe of the Ukraine is very favorable for the spring wheat yield, which is the same conclusion we drew for the Poltava experiment station, located in the southern part of the Ukrainian wooded steppe. It is possible that the drier autumn weather that characterizes the southern part of the Ukrainian steppe, as compared with the northern part

of the steppe, is also responsible for that importance of the autumn rainfall and temperature for the prerevolutionary peasant farm with its low level of cultural practice; but this inference in any case requires verification from data of the experiment stations located on the southwestern steppe part of the Ukraine with their own high levels of cultural practice.

The yield of spring wheat at the experiment stations, as comparison of the data shows, was considerably higher than the prerevolutionary peasant and estate yields.

This is explained by the considerably higher cultural practice at the experiment stations. As a general rule, spring wheat was included in a definite crop rotation and sown on land deeply plowed in the autumn (18-20 centimeters deep) it was sown early with a row drill, with a pre-sowing cultivation and a spring weeding of the sowings. All of this was responsible for a very significant increase of productivity on the fields of the experimental stations (for example, the series of yields taken by us at the Saratov station was more than double the prerevolutionary peasant and estate yield). But it must be admitted that the fluctuations about this considerably higher level still took place under the influence of the same natural forces. The studies of the dependence of the fluctuations in the prerevolutionary peasant yield, which were made in the arid zone, are enough to convince us of this. The work of A. M. Protasov on the Volga uplands of the former Saratov guberniya, which now belong to Saratov and Stalingrad oblasts, noted the very close connection of yield with the May-June rainfall and the saturation deficit and number of days with temperature of 30 degrees and over at 1300 hours for these two months.

The correlation coefficients for the combined action of these factors range from 0.77 to 0.88 (A. M. Protasov, The dependence between the yield of rye and wheat and the meteorological conditions in certain uyezds of the Lower Povolzh'ye on peasant farms. The work covers the period from 1892 to 1926). for the several Volga uyezds. An analogous study made by me on the peasant farms of the Trans-Volga uyezds of Novouzensk and Nikolayevsk, establishes a correlation coefficient of 0.917 between yield and the number of days with rainfall (1 millimeter and over) from sowing to ripening. Combining the index of saturation deficit at 1300 hours during the approximate period from tillering to the end of ripening increases this index to 0.955. (See the preceding article, and also the author's unpublished work "The dependence of yield on meteorological factors in the Trans-Volga area of the Middle and Lower Volga from 1886 to 1915". The estates with their higher mean productivity experienced the same yield fluctuations about that higher level and under the influence of the same causes. This has also been established by me in the above mentioned works.)

Thus one of the theses of the present work may be considered proved: that the yield fluctuations at the experiment stations, in spite of their high level of cultural practice, took place in the arid zone under the influence of the same natural forces that determined the upward and downward deviations of the prerevolutionary farm yields; but these fluctuations were against the background of the experiment stations' higher productivity, a circumstance which of course was of no little importance.

The use of snow retention as a special method of agricultural engineering, directed toward the increase of the soil moisture stocks by the beginning of the spring-summer period, considerably increases

the success of the struggle against drought, as was shown by the 18-year experience of the Saratov experiment station. Snow retention sharply increases the yields, which reached 14.61 tsentner per hectare (mean for the 18 years) as against 10.91 tsentner without snow retention, and also weakens the dependence of yield on the meteorological factors. The correlation coefficient between the yield with snow retention and the rainfall from germination to the end of ripening is + 0.729, while without snow retention for the same 18 years it is 0.018.

In the years of worst drought the plots with snow retention gave yields that were relatively "excellent" for those years. Even in the catastrophic year 1921 - a year of absolute drought - when less than 10 millimeters of rain fell at Saratov station during the entire growing season of spring wheat, the use of snow retention secured a harvest of 5.70 tsentner, equal to the mean prerevolutionary productivity in Saratov guberniya, while the spring wheat crop perished on the other plots. In 1924 and 1927, 4.35 and 5.73 tsentner per hectare respectively were harvested without snow retention at the Saratov experiment station. Snow retention increased the yield to 10.0 tsentner.

The effectiveness of snow retention is clear from the following table comparing the yield of spring wheat with and without snow retention during 18 years, arranged by varying amounts of rainfall.

TABLE 2

YIELD OF SPRING WHEAT IN TSENTNER PER HECTARE, WITH AND WITHOUT
SNOW RETENTION, ON THE FIELDS OF SARATOV EXPERIMENT STATION

Years grouped by amount of rainfall	Without snow retention	With snow retention
3 years with scantiest rainfall from germination to the end of ripening: 0-15 millimeters	3.72)	8.55)
	} 5.82	} 9.71
3 years with rainfall 16-65 mm.	7.93)	10.88)
3 years with rainfall 67-91 mm.	10.12)	14.01)
3 years with rainfall 99-109 mm.	13.07)	17.19)
3 years with rainfall 110-120 mm.	12.67)	17.10)
3 years with highest rainfall from germination to the end of ripening: over 120 mm.	15.20)	18.60)
	} 13.93	} 17.85

It appears from this table that yield increases with the rainfall on fields both without snow retention and with it, but that the yield with it exceeds the yield without it in all years.

Thus the use of snow retention in the arid zone makes it possible to increase the mean productivity of spring wheat to 15 tsentner per hectare (even without including that crop in a fertilized crop rotation) and weakens the dependence of yield on the elemental forces of nature. This great significance of snow retention has been taken into account by our socialist fatherland and this measure is extensively employed on kolchozes and sovkhoses.

Snow retention, however, only weakens the dependence of yield on weather, but does not liquidate it. The use of artificial irrigation, according to a work of V. Shneydyuller (V. Shneydyuller, Critical periods and the connection between the yield of winter and spring wheat under the conditions of dry and irrigated farming (from the data of the Kostychev experiment station) based on the materials of the 23-year experience of the Trans-Volga experiment station at Kostychev, leads, even with the crude methods of irrigation used at that station, firstly to an immense increase in mean yield (18.9 tsentner per hectare, against 5.70 tsentner on the control plots without irrigation), and secondly almost to the complete liquidation of the dependence of yield on the course of the weather, the correlation coefficient between the yield of irrigated spring wheat and the May-June rainfall being almost zero (+0.03), while for unirrigated spring wheat it is +0.0644. The harmful action of high temperatures and dry winds is also significantly attenuated (Table 3).

TABLE 3

CORRELATION COEFFICIENTS OF YIELD WITH HIGH TEMPERATURES AND
DRY WINDS AT THE KOSTYCHEV EXPERIMENT STATION

PLOTS	Number of days with harmful temperatures (30 degrees and over) in May and June	Dry-wind indices
Irrigated	- 0.400	- 0.184
Unirrigated	- 0.673	- 0.439

TABLE 4

YIELD OF SPRING WHEAT WITH AND WITHOUT IRRIGATION AT THE YERSHOV STATION

Plots	Yield of spring wheat in Tsentner per hectare on the Yershov tract, 1934
Unirrigated	10.5
Two irrigations, without fertilizer	24.0
One irrigation with fertilizer	25.2
Two irrigations with fertilizer	37.1

The experiments of agricultural stakhanovites have shown that even higher yields than this can be obtained. Thus, for example, V. G. Oleynikov, in the Stalingrad oblast', obtained 48-50 tsentner of spring wheat in the exceptionally arid year of 1936, on plowed-up virgin soil, with careful cultivation and weeding the sowing until all weeds had been completely destroyed, giving two irrigations. In the trans-Volga area of the Saratov oblast, V. I. Kursakov obtained a yield of 45.7 tsentner per hectare from a plot formerly sown to tobacco, with four snow retentions and two irrigations, using fertilizer (cf. "Agrotechnology of High Yields" published by VASKHNIL).

On generalizing the results of the studies at five experimental stations in the arid zone (Sinelnikovo, Saratov, Kostychev, Bezenchuk and Kupinsk), we arrive at the following conclusions:

(a) The higher level of cultural practice that allowed the experiment stations of the arid zone considerably to increase productivity did not free the yield of spring wheat from the influence of the meteorological factors, but the fluctuations took place about

a higher level of productivity. The principal factors involved in these fluctuations are the amounts of spring-summer rainfall, the spring-summer temperature regime, and the atmospheric humidity.

(b) The use of snow retention as a special supplementary method of agricultural engineering to increase the soil moisture stocks improves the position of spring wheat in the arid zone. Its mean productivity is sharply increased, and, as the 18-year experience of the Saratov experiment station shows, can be brought up to 15 tsentner per hectare, even in crop rotations without fertilizer, (The introduction of fertilizer into the crop rotations of the arid zone, or the use of special fertilizing for spring wheat, opens up the possibility of even further increases in productivity and its stability) on fields with snow retention, if superior cultural practices are used (drought-resistant varieties, autumn plowing, early row sowing, etc), ^{and} there is good management of the soil and the sowings. It also attenuates without destroying, the dependence of the yield on spring weather conditions, and even in years of catastrophic drought it saves the crops from total loss and guarantees relatively "excellent" yields for such years.

(c) Productivity in the arid zone is completely (or almost completely) liberated from the action of the elemental forces of nature under the conditions of irrigated farming. The system of irrigation revolutionizes the farming conditions of the arid zone, bring their yields to levels entirely unthinkable under the conditions of dry farming, while these harvests, freed from the influence of the caprices of the weather, acquire a stability which is of great importance for the national economy. A very high and stable yield - this is the result of the system of irrigated farming.

This system is already taking on reality on the fields of our socialist fatherland. On the initiative of the party, an extensive plan has been drawn for the irrigation of the Trans-Volga, the Manyok steppe, North Caucasus, Transcaucasia, Central Asia, etc., and this plan is already being accomplished. The time is not distant when "lean" harvests of spring wheat in the arid zone, even in the years of serious drought, will have passed into the realm of tradition.

2. The less arid wooded-steppes zone of the chernozem belt have been covered by investigations based on the materials of the following experiment stations: Chishma (Bashkir ASSR), Omsk (Omsk oblast), Poltava (southern part of the Ukrainian wooded steppe, and Sumy (northern part of the Ukrainian wooded steppe). The Balashov station of Saratov oblast, which is located in its more humid north-western part, should also be included in this category of stations.

First of all we note one feature common to all these five experiment stations: the period from sowing to germination is characterized by small negative correlation coefficients between yield and rainfall, which testifies to the inverse connection of the spring wheat yield with the rainfall during this initial phase of plant life. The correlation coefficient ranges from -0.103 at Chishma to -0.398 at Omsk. Rainfall in regions not extremely arid apparently packs the soil and increases its moisture content, thus interfering with the processes of soil aeration which are so essential for the germinating grain. In the arid zones this negative effect of abundant rainfall is overshadowed by its importance for the future development of the plants, since it also increases the soil moisture stocks. Early drought in the period from sowing to germination is of course more frequent here than in regions with a more humid climate. This is

why the period from sowing to germination both in the eastern Ukraine and on the Trans-Volga steppes (Sinel'nikovo, Saratov, Besenchuk) (the correlation coefficient of yield with the number of humid days is $+0.214$ at Saratov, $+0.533$ at Besenchuk (rainfall), $+0.630$ at Sinel'nikovo (rainfall)). The matter is apparently different at Kipinsk experiment station, where the first two May decades give a coefficient of correlation with a minus sign, between yield and rainfall. Under the conditions of the harsh Siberian climate, excessive moistening of the soil at the onset of spring retards its warming up after the winter season and acts injuriously even in the steppe regions of Siberia, while on the wooded steppe (Omak) the harmful influence of rainfall from sowing to germination, peculiar to wooded steppes in general, is added to this. The fact that the Omak experiment station has the highest correlation coefficient with a minus sign (-0.398) is apparently explained by this dual action) is characterized by positive correlation coefficients, while in the regions of milder climate with respect to drought conditions (Balashov, Poltava, Sumy, Chistom, Omak), low negative values are prevalent. Moreover, at Balashov experiment station, curvilinearity could be noted in the reaction of yield to the rainfall from sowing to germination; the yield is impaired both by excessively dry weather and excessive moisture, but in general the negative influence of the latter is greater. Somewhat dryish weather from sowing to germination is the most favorable for the yield in the region of the Balashov station.

Thus the first distinction of the wooded-steppe belt, and, in general, of regions with a milder climate, from the arid zone, consists in the transformation of the positive effect of rainfall to a weakly negative action during the initial period of the development

of spring wheat. A second distinction consists of the attenuation of the influence of rainfall, high temperatures, and saturation deficits in the succeeding periods, when these elements of soil and atmospheric drought act in the same direction as in the arid zone. Comparison of the stations with the mildest climate and the stations in the arid zone nearest them will convince us of this.

We begin with the eastern stations:

(a) Kupinsk and Omsk experiment stations. The correlation coefficient of yield with the rainfall in June and the first July decade is + 0.90 at Kupinsk experiment station, while in the wooded steppe station of Omsk, the yield correlates with rainfall from germination to full heading to the extent of + 0.415, and with rainfall from full heading to waxy ripeness /ВОСКОВОЙ СПЕЛОСТИ/ to the extent of + 0.224. No explanation at all is needed to see that the difference in these correlation coefficients indicates the great weakening in the influence of the rainfall on the yield in the wooded steppes of Siberia. Even though the comparison is not entirely above reproach, since we are comparing the vegetative periods of one station with the calendar periods of another, (the material for conversion of the meteorological indices of the Kupinsk station to a vegetative period basis is not available) still, since the dependences at Omsk and Chisima are almost identical (the data are presented in unpublished studies of these stations by the author), the Omsk station has been included in the comparison with Bezenduk on the same basis with that of Chisima.

TABLE 5

COMPARATIVE TABLE OF CORRELATION COEFFICIENTS OF SPRING WHEAT YIELD WITH RAINFALL AND TEMPERATURE AT BEZENCHUK,
CHISHMA AND OMSK EXPERIMENT STATIONS.

Vegetative periods (I have no detailed phenological data for Chishma and Omsk stations. The material received divides the entire growing season for each year into three phases: sowing to germination, germination to full heading, full heading to waxy ripeness. The onset of blooming almost coincides with full heading)

1. Sowing to germination
2. Germination to onset of blooming (full heading)
3. Onset of blooming (full heading) to maturation

Correlation coefficients of the yield of spring wheat					
with the rainfall			with the temperature at 1300 hours		
Bezenchuk	Chishma	Omsk	Bezenchuk	Chishma	Omsk
+ 0.533	- 0.103	-0.398	-	-	-
+ 0.891	+ 0.423	+0.415	- 0.569	- 0.498	- 0.478
+ 0.138	+ 0.121	+0.224	- 0.461	- 0.433	- 0.523

(b) Bezenchuk, Chishma and Omsk stations. Still considering the first period (sowing to germination, see Table 5), which has already been mentioned, we see that a fundamental quantitative distinction between the experiment station at arid Bezenchuk and the other two stations of the eastern wooded steppe is observed throughout the entire growing season, down to the onset of bearing. The correlation between the yield at Chishma and Omsk and the rainfall of this season is considerably lower than at Bezenchuk, which is located in the arid zone. The third period, that of grain formation, is marked by very low correlation coefficients at all three stations. In exactly the same way there is apparently no substantial difference between the reaction of yield to the temperature conditions of the second and third periods: the correlation coefficients are approximately of the same order for all three stations, and all of them have the minus sign, which shows an inverse relation between yield and temperature (cf. Table 5).

(c) Saratov and Balashov experiment stations. A special processing of the Balashov data showed that under the conditions at this station the positive influence of rainfall on the spring wheat yield is expressed in the period from germination to the onset of ripening, including the blooming phase, with a correlation coefficient of + 0.619. This period from germination to the onset of ripening (to render the comparison with the Balashov station more precise, the Saratov data was converted to show the distribution among the phases in which the positive influence of rainfall and the negative influence of high temperatures are most effective for Balashov) is marked at the Saratov station by a considerably closer relation between yield and rainfall: the coefficient of correlation rises here to + 0.819. This difference is even sharper for the tem-

perature at 1300 hours, from the onset of tillering to the onset of ripening; the correlation coefficient for Balashov is -0.440 , but at Saratov it is -0.317 .

There is thus undoubtedly an attenuation of the positive influence of rainfall and the negative influence of high temperatures on the spring wheat yield at Balashov station, which is located in an area with milder climate.

We note that the final period from the onset of ripening to harvest time is marked at Balashov by a curvilinear correlation of yield with the rainfall of that period; both absence and abundance of rainfall injure the yield, but the negative influence of high yields prevails, which fact is expressed by low correlation coefficients, with the minus sign, between yield and the rainfall from the onset of ripening to harvest time. The same curvilinearity, and for the same period, is also observed for the temperature conditions: moderately hot weather from ripening to harvesting is the most favorable for the yield of spring wheat at Balashov station.

The curvilinear character of the dependence of the spring wheat yield at Balashov station, from the onset of ripening down to the harvest time, is apparent from the 23-year observations on rainfall:

MEAN YIELD IN TSENTNER

PER HECTARE

6 years with scantiest rainfall (4-13 millimeters)	9.9
6 years with rainfall 27-27 millimeters	10.9
6 years with rainfall 28-71 millimeters	10.6
5 years with heaviest rainfall (90-137 millimeters)	8.3

Mean Yield in Tsentner

Per Hectare

By the temperature readings at 1300 hours:

6 years with lowest mean temperature at 1300 hours: 22.4 degrees	8.3
6 years with next highest temperature: 24 degrees	7.9
6 years with next highest temperature: 25.9 degrees	14.7
5 years with highest mean temperature: 28.6 degrees	8.7

We remark that a certain curvilinearity in the dependence of yield on rainfall is also observed for the period from the beginning of sowing to the emergence of the seedlings. Rather dryish weather during this period, with scanty rainfall, gives the best results, as may be seen from the following table:

Mean Yield in Tsentner

Per Hectare

6 years with scantiest rainfall: 0-2.4 mm.	8.8
6 years with rainfall 2.6 - 8 mm.	13.2
6 years with rainfall 9.8 - 20 mm.	10.2
6 years with heaviest rainfall (over 20 mm.)	7.6

The yield is harmed by both excessive and low temperature. This curvilinearity in the reaction of yield to rainfall and temperature is encountered even more strongly at the Poltava experiment station in the southern part of the Ukrainian wooded steppes.

(d) Sinel'nikovo and Poltava experiment stations. A

comparison of the Poltava station with the arid steppe station of Sinel'nikovo confirms the conclusions drawn for the eastern zone (see Table 6)

A direct relation of the yield with rainfall and an inverse relation with temperature at 1300 hours are observed for all periods at Sinel'nikovo, while at Poltava a positive influence of rainfall - in attenuated form at that - is noted only for the phases from germination to blooming.

[See next page for Table 6]

The initial and final periods in the life of spring wheat are marked by the negative effect of rainfall on the yield. For temperature we note an insignificant positive action from sowing to tillering time at Poltava, the absence of any dominating influence at all, whether positive or negative, during the grain-formation phase at that station, while at Sinel'nikovo a fairly marked inverse relation of yield to temperature is observed for these periods. The tillering to blooming period also shows an inverse relation at Poltava, but to a lesser extent than at Sinel'nikovo.

There are not all the peculiarities of the Poltava station, however: there is a definitely curvilinear correlation, during the period from tillering to blooming, of yield with rainfall and temperature, and the points of optimum rainfall and temperature are distinctly expressed. The optimum rainfall in this period is 130 millimeters, and optimum 1300 hours temperature in this period is 19.6 degrees. On both sides of these points the yield declines. Thus under the wooded-steppe conditions of the Poltava station, cases of the harmful effect of abundant rainfall and low temperatures during

TABLE 6

CORRELATION COEFFICIENTS OF YIELD WITH RAINFALL AND TEMPERATURES AT SINEI'NIKOV AND POLTAVA EXPERIMENT STATIONS

PHASES OF DEVELOPMENT ¹	CORRELATION WITH RAINFALL		COEFFICIENTS OF YIELD WITH TEMPERATURE AT 1300 HOURS	
	SINEI'NIKOV	POLTAVA	SINEI'NIKOV	POLTAVA
Sowing to onset of germination	+ 0.630	-0.172)	-0.584	+0.186
Germination to onset of tillering	+ 0.234	+0.182)		
Tillering to onset of blooming	+ 0.747	+0.436	-0.505	-0.401
Blooming to maturation	+ 0.498	-0.341	-0.606	-0.024

¹ The dates of the onset of tillering are given at Poltava station with great gaps: not for all years. Only approximate calculations have therefore been made for Poltava (the details are set forth in the unpublished working-up of the author for Poltava experiment station.

this period are observed, though in general the positive influence of rainfall and the negative action of high temperatures predominate. When the curvilinearity of the reaction of yield to rainfall and temperature conditions is taken into account, the coefficient of correlation rises from 0.436 (rainfall) to 0.526, and from 0.401 (temperature) to 0.517.

The same curvilinear reaction of yield to temperature is also noted for the period from the onset of blooming to ripening (details in the unpublished work-up for Paltava station). At the first glance, the temperature of this phase, with a correlation coefficient of -0.012, has no influence on the yield. But trials with curvilinearity revealed an entirely definite parabolic dependence, with an optimum temperature point almost coinciding with the long-term mean temperature for this period, and with equally developed branches of the parabola characterizing the deterioration of grain-formation conditions with both falling and rising temperatures. When "curvilinearity" is taken into account, the correlation coefficient rises from zero to 0.306. The relation between the yield of spring wheat and the rainfall of this period (onset of blooming to ripening) is inverse and is expressed by the negative correlation coefficient -0.410.

Finally, analysis of the autumn weather conditions in the year preceding the harvest allowed us to establish a relation, between yield and late autumn rainfall, measured by the correlation coefficient + 0.519. Tests for curvilinearity gave positive results: the correlation coefficient rises very appreciably to reach 0.665. The nature of the relation between yield and late autumn rainfall is different, however, than in the cases considered above. No

points of optimum rainfall were observed at all. The curve has the form of a parabola, with accelerating growth of the yield, within the field of observation, as the rainfall increases.

Scanty and average rainfall have almost identical effects on the yield, and only abundant rainfall in late autumn, which deeply penetrates into the soil, creates such moisture stocks in it as possess essential significance for the spring vegetation of spring wheat. An early, cool and long protracted autumn, with a large amount of rainfall during it, are the conditions that favor the future yield of spring wheat at the Poltava experiment station. Of the 7 years with the highest late-autumn rainfall, 5 had over-average yields (18.2 to 29.4 tsentner per hectare), while 2 were near the average (14.1 - 14.3 tsentner), the average 36-year yield being (1887, 1895 and 1920, with 29.4, 27.7 and 22.2 tsentner respectively) belong to the group of these 7 years with abundant late-autumn rainfall.

Since principal factors that influence the yield of spring wheat are thus noted at the Poltava experiment station: (1) late-autumn rainfall; (2) and (3) rainfall, temperature and rainfall during the blooming phase. The correlation coefficient of their combined action, taking the "curvilinearity" of the influence on yield into account, attains the value of 0.800 (the details are set forth in an unpublished study of the Poltava experiment station by the author; see also Diagram 2).

DIAGRAM 2.

YIELD OF SPRING WHEAT AT POLTAVA EXPERIMENT STATION

YIELD IN TONNES PER HECTARE

Observed yield

Yield calculated from the regression equation

$$\begin{aligned}
 \bar{y} &= -22.2657 + 0.8053 y + 0.4154 y \\
 &+ 0.4544 + 0.3439y + 0.4538y ; \\
 R &= 0.822 ; R_1 = 0.796
 \end{aligned}$$

But the arid Sinel'nikovo experiment station obtained the same result (a correlation coefficient of 0.802) by including only two factors in the equation (rainfall from tillering to blooming, and saturation deficit at 1300 hours from blooming to maturation); while the combined influence of three factors (the percentage of days with useful rains (a day is considered to have useful rain if the rain gage reading exceeds the evaporation as measured by the evaporimeter) separately for three periods: (1) sowing to tillering, (2) tillering to blooming, (3) blooming to maturation) exceeds the action of five factors for Poltava, even when "curvilinearity" is taken into account, and is expressed by the correlation coefficient 0.917 (see the unpublished material on the Sinel-nikovo station). All this argues for the dispersal of the influence of the meteorological factors at Poltava, as compared with Sinel'nikovo; here there is no predominant influence exerted by one or two factors, but the combined influence of these five factors, which do not always work in the same direction at that, acquires the dominating significance in the fluctuations of the spring wheat yield on the fields of the Poltava experiment station (with the correlation coefficient 0.796). It subsequently proved possible to establish the appreciable influence

of the action of insect pests on the Poltava yield of spring wheat (mainly the Hessian fly, then the Swedish fly and other pests) and of fungous diseases (mainly rust, smut, etc). For the 20-year interval for which material is available (1887-1906), the damage done to the crop by insects and fungus was very considerable in 1894 and 1897; while in six years (1889, 1893, 1895, 1896, 1902, 1906) the damage was serious. In general the remaining years were good in this respect (unpublished material on Poltava experiment station). Table 7 shows the effect of pests and fungi on the yield of spring wheat.

TABLE 7

INFLUENCE OF PESTS AND FUNGUS DISEASES ON THE YIELD OF SPRING WHEAT
(POLTAVA EXPERIMENT STATION)

Groups of Years	Mean yield in tsentrner per hectare	
1st group. 2 years with very serious damage	10.6	
2nd group. 6 years with serious damage	14.4	13.2
3rd. group. 12 years with no damage	16.6	

The correlation coefficient between the yield of spring wheat and the degree of damage by insects and fungi is 0.368 (the following numerical values for the degree of insect and fungus damage were correlated with the yield: 1. very serious; 2. serious damage; 3. Favored years. (no damage)). If this factor is included in the regression equation to supplement the meteorological factors already mentioned, the correlation coefficient reaches the value of 0.935. The coefficient of partial correlation between the yield and the degree of damage by insects and fungous diseases, eliminating the meteorological

factors, is 0.654, while the coefficient of partial correlation of the yield, eliminating insect and fungous damage, is 0.924. Thus the influence of meteorological factors, eliminating the harmful consequences of the action of these biological agents, stands out more distinctly and in sharper relief, just as when we eliminate the effects of the meteorological factors, the results of the damage done to the spring wheat by pests and fungous diseases is outlined more clearly and brightly. The correlation between the yield and the meteorological factors included in this study is expressed in this case by a value equal to 0.931, i. e. it reaches the general level fixed for three factors at the Sinel'nikovo station (0.917).

We had two reasons for dwelling on the computation of the influence of insect and fungus damage on the yield at Poltava station: first, because the Poltava experiment station is our only station with 20-year systematic observations of the damage caused to crops by insects and fungous diseases; (these data are taken from K. G. Man'kovskiy's work "Results of twenty years' work on the Poltava experimental field (1866-1906), fascicle II. A question on damage from insects and fungous diseases was included in the questionnaire that was sent to the experiment stations for the study on "Climate and Wheat". But the answers to this question were only fragmentary. Sinel'nikovo station indicated that rust did serious damage to spring wheat in 1930. Omsk notes the same for 1928. The Semy station mentions the damage caused by the Hessian fly to spring wheat in 1925, 1926, 1931, and especially in 1932) and second, also because a peculiarity distinguishing the wooded steppe of the Ukraine from its steppe portion affects this situation. The Ukrainian wooded steppe is a region in which the Hessian fly is disseminated,

and the fungous diseases, especially rust, attack plants oftener and more strongly in the more humid climate of the wooded steppe than in the steppe zone.

(c) The Poltava and Sumy experiment stations. Let us compare the results of the study on the Sumy station, in the North of the Ukrainian wooded steppe (at the border of Polesia) and the southern wooded-steppe station at Poltava. The variations in the reaction of the yield to meteorological factors also correspond to the climatic peculiarities of these two stations.

Let us begin our comparison with the rainfall. The late-autumn rainfall exerts a positive influence on the yield at the Sumy station, but in a considerably attenuated form; the predominant role of the heavy rains is absent here: the correlation coefficients here is + 0.297, and tests for curvilinearity did not give positive results.

At the Poltava experiment station, as we shall see, the relation between yield and late autumn rainfall is more strongly expressed, with a special place reserved for the abundant rains that deeply soak the soil (the correlation coefficient for Poltava is + 0.52, or + 0.65 after taking account of curvilinearity, which characterizes the role of the heavy rains.) This difference in the quantitative influence of the autumn rainfall has a real foundation: the climate is drier in the region of Poltava station, which consequently needs rainfall more than Sumy station, which already feels on it the "breath" of Polesia. The similarity and differences in the effect of the spring rainfall on the yield of spring wheat are explained by the same reason. (It is "wooded steppe" in both cases, but Poltava is southern wooded steppe, while

Suny is northern).

At Poltava station, the rainfall of the period 30-40 days before and up to blooming, which has a clearly expressed curvilinearity, is of the most importance for the yield, in the entire course of the spring-summer period; but all the same, the positive influence of rainfall is in general predominant (the correlation coefficient is $+0.43$, or $+0.52$ with the curvilinearity). At the first glance, the positive influence of rainfall on the spring wheat yield is almost absent at Suny station (the details are set forth in a special unpublished work-up of the data of that station). All variants of the combination of the elementary short /vegetative/ periods, each with negligible correlation coefficients, lead to coefficients of the order of $+0.106$, $+0.135$. Tests for curvilinearity however, reveal a higher value for it: the correlation coefficient for the rainfall from germination to the onset of heading rises from $+0.106$ to 0.62 , which is evidence that at Suny both insufficient and excess rain have a negative influence: in both cases the yield of spring wheat is injuriously affected. Comparison with Poltava indicates that the curvilinearity of the reaction of yield to rainfall is far more sharply expressed at Suny.

With reference to the first period, from sowing to germination, a similarity is observed between the stations. In both cases the rainfall exercises a weak and dominantly negative influence on the yield of spring wheat (Poltava, -0.17 ; Suny, -0.24). Table 6 gives a clear idea of how the difference in the climatic conditions of Poltava and Suny is reflected in the reaction of yield to temperature.

As we see, the first two periods give almost identical correlation coefficients, which points to the weakly dominant positive

influence of the temperature from sowing to tillering, and to the appreciably greater and again dominant negative influence of high temperatures from tillering to blooming. But the final phase, that of the grain-formation, marks an essential difference in the action of the temperature conditions on the yield. Poltava is characterized by a zero coefficient of correlation. As we have seen, a definite and clearly expressed curvilinearity in the reaction of yield to temperature is observed at this station; both low and high temperatures injure the yield. Under the climatic conditions of Sumy, the predominance of the positive influence of temperature is unconditionally clear.

Thus, passing from the arid steppes of the Ukraine (Sinel'nikovo) through the southern wooded steppes (Poltava) to the northern wooded steppes (Sumy), we witness the successive transformations of the sharply dominant negative influence of high temperatures during the period of grain formation (Sinel'nikovo, -0.636), through the curvilinear character of the dependence of yield on temperature conditions, where there is no dominantly positive or negative reaction of yield to temperature (Poltava: linear correlation coefficient is -0.014 ; parabolic, from flowering to ripening, is -0.306), to the dominating positive influence of warmth (Sumy, $+0.385$).

[See next page for Table 8]

For completeness of comparison between the two "contrast" stations, Sinel'nikovo and Sumy, we compare in Table 9 the basic characteristics of the influence of meteorological factors on the yield during the several vegetative periods.

At Sumy, we see the marked attenuation of the positive

TABLE 8

INFLUENCE OF TEMPERATURE ON THE YIELD OF SPRING WHEAT AT POLTAVA AND
SUMY EXPERIMENT STATIONS

Periods	Coefficients of correlation between yield and temperature at 1300 hours	
	POLTAVA	SUMY
From sowing to tillering (the onset of blooming at Sumy have been calculated approxi- mately)	0.136	0.216
Tillering to blooming	-0.401	-0.159
Blooming to maturation	-0.014	0.305

influence of the rainfall and the negative influence of temperature, and the appearance of the harmful effect of excessive moistening and reduced temperatures. At Sinel'nikovo, the rainfall exerts a positive influence on the yield in all periods. The correlation coefficients have the plus sign and are large enough in size; while at Sumy the harmful effects of insufficient moistening are discovered only when the curvilinearity of the reaction of yield to rainfall from the emergency of the seedlings to blooming is taken into account: either insufficiency or abundance of rain during this period damage the yield. Sinel'nikovo station is characterized by an inverse

TABLE 3

INFLUENCE OF TEMPERATURE ON THE YIELD OF SPRING WHEAT AT POLTAVA AND

SUNY EXPERIMENT STATIONS

Periods Coefficients of correlation between yield and
temperature at 1300 hours

POLTAVA

SUNY

From sowing to
tillering (the
onset of blooming
at Suny have been
calculated approxi-
mately)

0.186

0.216

Tillering to
blooming

-0.401

-0.159

Blooming to
maturation

-0.014

0.385

influence of the rainfall and the negative influence of temperature, and the appearance of the harmful effect of excessive moistening and reduced temperatures. At Sinel'nikovo, the rainfall exerts a positive influence on the yield in all periods. The correlation coefficients have the plus sign and are large enough in size; while at Suny the harmful effects of insufficient moistening are discovered only when the curvilinearity of the reaction of yield to rainfall from the emergency of the seedlings to blooming is taken into account: either insufficiency or abundance of rain during this period damage the yield. Sinel'nikovo station is characterized by an inverse

relation between yield and temperature for all periods of the spring wheat growing season.

[See next page for Table 2]

At Sumy station the negative influence of high temperatures prevails only during the middle period from tillering to blooming, while in the extreme periods of the life of spring wheat, the positive influence of warmth is dominant.

We note the following NEVANKI lacunae for Sumy station:

1. The period from the onset of tillering to blooming is characterized by a negative correlation coefficient for temperature; the following period, that of grain formation, is characterized on the contrary by its positive action. The misalignment consists of the fact that curvilinearity, as with the Poltava station, could not here be found in the reaction of the yield, in the former case, to the predominantly negative influence of temperature, and in the latter case, to its positive effect. This curvilinearity apparently also holds for Sumy station, but could not be discovered in the short, 16-year series of observations. For this reason I prefer to speak simply of the prevalent influence of the temperatures during these periods.

2. There is a more substantial lacuna in the rainfall of the bearing period. For Poltava, the correlation coefficient between the yield and the rainfall for this period has a minus sign. One would think that at the more "humid" Sumy station this minus sign, indicating the negative influence of the rainfall during the bearing period, would still remain. But we have here instead is a

TABLE 9

CORRELATION COEFFICIENTS OF YIELD WITH RAINFALL AND TEMPERATURE AT 1300 HOURS AT SINEL'NIKOV AND SUMY STATIONS

PERIODS ¹	CORRELATION COEFFICIENTS OF YIELD WITH RAINFALL		PERIODS ¹	CORRELATION COEFFICIENTS OF YIELD WITH TEMPERATURE AT 1300 HOURS	
	SINEL'NIKOV	SUMY		SINEL'NIKOV	SUMY
Sowing to germination	+ 0.630	-0.243	Sowing to tillering	-0.584	+0.216
Germination to blooming	+ 0.700	+0.106	Tillering to blooming	-0.505	-0.459
		(with curvilinear- linearity)			
Blooming to maturation	+ 0.497	+0.056	Blooming to maturation	-0.606	+0.385

¹ (The division into vegetative periods has been made for the sake of full comparability of the comparisons under the peculiar scheme set up for the Sumy station, which has proved the most effective for it (set forth in the Sumy station materials).)

correlation coefficient almost equal to zero ($+0.056$), while no curvilinearity in the reaction of yield to rainfall could be found. Obviously this period requires supplementary and more detailed analysis. We assume in general that the period of ismel formation (from blooming to maturation) can be completely elucidated only through the parallel treatment of the data on the yield of straw and that of grain. The complexity of this parallel analysis of the yields of straw and grain made it impossible to realize this intention. Specifically, too, it could not be carried out for Sumy station, since the materials it sent did not include straw-yield data.

To generalize all that has been said about the influence of meteorological factors on the fluctuations of the spring wheat yield in the wooded-steppe zone, we see that both in its eastern and western (Ukrainian) regions an attenuation is in general observed in the effect of drought and high temperatures on the yield, in comparison to the steppe zone. This phenomenon, however, is more strongly expressed on the western Ukrainian wooded-steppe. Specifically, the bearing period at the eastern stations (Chischna, Zuck) is still marked by appreciable negative action of high temperatures on the yield of spring wheat, while at Poltava experiment station the influence of the temperature conditions of this period has a dual character: both low and high temperatures harm the yield, while, at the northern wooded-steppe station of Sumy, the positive influence of warmth is already prevalent.

2. WINTER WHEAT

1. Throughout the zone of the most widespread distribution of winter wheat, we note first of all that everywhere the autumn rain-

fall of the year before the harvest exerts a positive influence. This dependence is observed at the stations of the Azov-Black Sea Kray (Yeyak, Rostov, Krasnodar) as on the fields of the steppe and wooded steppe Ukrainian experimental institutions (Kherson, Odessa, Sinel'nikovo, Poltava and Sumy experiment stations), and is noted at the Shatilov station of Orel oblast. A more or less marked influence of the autumn rainfall should apparently be attributed to the southern steppe belt; at any rate the highest correlation coefficients are found at the southern stations (Rostov, + 0.639; Krasnodar, + 0.430; Odessa, + 0.52) (The conclusions of the Azov-Black Sea Kray are based on the unpublished work of A. P. Prolov "Influence of meteorological factors on the yield of 'Ukrainka' winter wheat under the conditions of the Azov-Black Sea Kray". In view of the fact that the author of that work used the not irreproachable method of combining the data for different stations in a single series, I have segregated the three stations (Yeyak, Krasnodar, Rostov-on-Don) and have worked out the data separately for each of them, covering the autumn and winter periods. The conclusions of my analysis confirm those of Prolov. The spring-summer period for Rostov was worked out independently by me. Since all the necessary information for the study of the spring period at Yeyak and Krasnodar was lacking, this study had to be given up. For Sinel'nikovo experiment station the work of V. A. Ravich "The yield of winter wheat in relation to the principal weather elements" was used. This work was published at Dnepropetrovsk in 1928). These coefficients for the other stations all have the plus sign, but do not go above 0.378. The autumn period, however, does not play a major role in the fluctuations of the yield of winter wheat.

2. The wintering conditions exert a considerably more powerful influence, which is likewise found everywhere. At the three Azov-Black Sea stations, the correlation coefficients of the yield with the aggregate of the negative minimum temperatures during the winter period notably exceed the correlation coefficients between the yield of winter wheat and the autumn rainfall (Table 10)

TABLE 10.

CORRELATION COEFFICIENTS OF THE YIELD OF WINTER WHEAT WITH THE
RAINFALL AND TEMPERATURE

	COEFFICIENTS OF CORRELATION		
	ROSTOV	YEYSK	KRASNOVAR
With the rainfall	+ 0.639	+ 0.084	+ 0.430
With the aggregate of the negative minimum temperatures during winter, late autumn and early spring	+ 0.856	+ 0.691	+ 0.551

At Sinsl'nikovo station, in the course of an 11-year period, the winter wheat was completely destroyed by freezing on three occasions, in one year it suffered badly and yielded the very low crop of 11.04 tsentner per hectare, while in the other 7 years, with good wintering conditions, its mean yield was 20 tsentner per hectare, with the lowest, but still fairly good yield of 14.05 tsentner in 1933. (This lowest yield of the years with favorable wintering conditions exceeds the mean 11-year yield, which was 13.8 tsentner per hectare).

At the Sumy station the average yield of wheat over 22 years of observations was 15.8 tsentner per hectare. In three cases (1920, 1929, and 1932), the wheat crop was killed by freezing, with rather noticeable effects on the mean productivity. The mean yield for the 19 years without winterkilling of the sowings was 18.2 tsentner, and for the entire 22 years it was 15.8 tsentner. This does not, however, exhaust the enumeration of the influences exerted by the wintering conditions on the yield fluctuations of winter wheat, for even in the years without winterkilling, there were still cases of damage to the winter wheat crop by unfavorable wintering conditions, which damage ranged from very serious to insignificant. But for Sumy station this information is not available to us, and only the years of total crop loss are given. The only station for which I have detailed data of the winter damage for 21 years, from 1886 to 1906, is that of Poltava, in the south of the wooded steppe. For this reason the materials of this station were worked up in detail (the author's study of the Poltava station is unpublished). In 1891, at Poltava station, the winter wheat crop was completely destroyed by an extremely severe winter. All of it was killed on each plot, of the experimental field regardless of variety or cultural practice employed. But this was the only case in the 21 years when the wheat crop was totally lost. The remaining 20 years are distributed as follows, by degree of winter damage:

YIELD IN TSENTNER PER HECTARE

2 years with very serious winter damage	6.7
12 years with serious winter damage	13.2
2 years with appreciable winter damage	19.5

YIELD IN TSENTNER PER HECTARE

3 years with appreciable winter damage	21.4	22.3
9 years with entirely fortunate wintering conditions	22.6	
Mean 20-year yield	18.9	

As we see, the range in the yield of winter wheat, according to the degree of winter damage, is very considerable. Of the 12 more or less fortunate years, only two gave reduced yields: 1894, when the crop suffered badly from rust, with a yield of 13.5 tsentner per hectare, and 1888, with 18.3 tsentner per hectare, when the yield of winter wheat was considerably reduced by its lodging on the fertilized plots.

Using the method of calculating the correlations, the influence of the wintering conditions on the yield can be more definitely expressed: the correlation coefficient between yield and the degree of winter damage is 0.799. (The yield was correlated with the numerical values: 1, very serious damage; 2, serious damage; 3, appreciable damage; years with insignificant damage or none at all designated with 4 balls). The size of this coefficient allows the following conclusion to be drawn: at the Poltava experiment station, located in the southern part of the Ukrainian wooded steppe, on the boundary of the steppe, the wintering conditions for winter wheat are the predominant factor influencing the fluctuations in the yield of that crop.

It is clear from the description of the winter weather in the "Results of the work at Poltava experimental field for 20 years

(1886-1905)", fascicle II, that the unfavorable wintering conditions might be the result of a wide range of meteorological situations. Alongside of the harsh winters with their deadly effect on winter wheat (for example, 1890/91) we also meet "muggy" winters, with thaws, with alternate freezing and thawing of the soil, with ice crusts, water standing in the fields, leading to the VYPREVANIYE or snow-suffocation of winter wheat. Finally, in still other years, the period of transition from winter to spring likewise caused no little damage to winter wheat from the abrupt changes of temperature.

We have unfortunately no such detailed annual characterizations from the other stations of the Ukraine and Azov-Black Sea Kray - the more so since the relatively prolonged period of 21 years is involved - and can therefore not assert that the yield fluctuations of winter wheat are determined to the same extent by the wintering conditions at such other stations as is observed at Poltava. In any case the data given above for the Rostov, Yaysk, Krasnodar and Sinel'nikovo stations speak of the considerable influence of winter weather on the yield of winter wheat both in the steppe belt of the Ukraine and in the Azov-Black sea area, but the relative importance of the wintering conditions is apparently less in the south than on the wooded steppe, since the influence of autumn and spring weather has a stronger effect on the arid steppe than in the more humid climate of the wooded steppe. We have already spoken of the more appreciable positive significance of the autumn rainfall in the southern belt; the same is observed with respect to the spring rainfall, to which we now pass.

3. The work of V. A. Ravich on winter wheat at Sinel'nikovo experiment station notes the positive influence of the May-June rainfall on the yield, expressed in the correlation coefficient of + 0.55, and the strong negative action of the May-June saturation

deficit, with a correlation coefficient of -0.67 . The positive effect of the May rainfall and the harmful influence of high May temperatures is noted by N. I. Sofoterov, for the Odessa and Kherson stations, with correlation coefficients at Odessa of $+0.52$ and -0.52 respectively. In the Azov-Black Sea Kray, under the conditions of higher cultural practice, the influence of rainfall from the onset of spring to and including full heading is apparently absent (at least for the "Ukrainka" variety), but the rather appreciable positive influence of rainfall during the ripening period is observed there, although undoubtedly the most powerful effect is that of the dominantly harmful influence of high temperatures from the onset of spring to maturation, as shown by analysis of the data for Rostov station. The correlation coefficients between the yield of winter wheat and the temperature at 1300 hours at the Rostov station are as follows:

Four early spring decades	-0.717
The three decades before heading	-0.394
The three decades after the onset of heading	-0.590

Atmospheric drought, of which high temperature constitutes an index, thus unconditionally exerts a great and harmful influence under the conditions of the Rostov station.

Thus the spring-summer weather on the steppe portion of the Ukraine and in the Azov-Black Sea Kray has an influence on the winter wheat yield in varying form and degree, but in any case a quite appreciable influence, alongside of the autumn weather and wintering conditions, while apparently, too, a principal (though

not a dominant) part in this "competition" of meteorological factors belongs to the wintering conditions (see the data given above for Rostov, Yeysk, Krasnodar and Sinel'nikovo).

A different picture is observed at the southern wooded-steppe station of Poltava, where the climate is more humid. Here the wintering conditions do dominate, meeting, as they do, only weak competition from the meteorological factors of the autumn and spring-summer periods. The correlation coefficients between the winter wheat yield at Poltava station and the rainfall in various vegetative periods are as follows:

Autumn rainfall	+ 0.171	+ 0.256
Spring rainfall, to blooming	+ 0.181	
From blooming to harvest time	- 0.389	

The coefficients of correlation of yield with the autumn and spring rainfall have the plus sign, but are of negligible size. The rainfall during the period of grain formation have a stronger influence on the winter wheat yield under the conditions of the Poltava station, but the minus before the correlation coefficient indicates its negative influence on the yield. The combined rainfall influence, however (positive for autumn and spring, negative for the grain-formation phase) is measured by a correlation coefficient of + 0.432 (established in an unpublished study of Poltava station by the author). Thus, on one hand, the influence of wintering conditions here is measured by the correlation coefficient 0.799, while on the other hand the effect of the autumn, spring and summer rainfall together is characterized by the correlation coefficient 0.432. This comparison needs no explanation to indicate the

not a dominant) part in this "competition" of meteorological factors belongs to the wintering conditions (see the data given above for Kostov, Yoyak, Krasnodar and Sinel'nikovo).

A different picture is observed at the southern wooded-steppe station of Poltava, where the climate is more humid. Here the wintering conditions do dominate, meeting, as they do, only weak competition from the meteorological factors of the autumn and spring-summer periods. The correlation coefficients between the winter wheat yield at Poltava station and the rainfall in various vegetative periods are as follows:

Autumn rainfall	+ 0.171	+ 0.256
Spring rainfall, to blooming	+ 0.181	
From blooming to harvest time	- 0.389	

The coefficients of correlation of yield with the autumn and spring rainfall have the plus sign, but are of negligible size. The rainfall during the period of grain formation have a stronger influence on the winter wheat yield under the conditions of the Poltava station, but the minus before the correlation coefficient indicates its negative influence on the yield. The combined rainfall influence, however (positive for autumn and spring, negative for the grain-formation phase) is measured by a correlation coefficient of + 0.432 (established in an unpublished study of Poltava station by the author). Thus, on one hand, the influence of wintering conditions here is measured by the correlation coefficient 0.799, while on the other hand the effect of the autumn, spring and summer rainfall together is characterized by the correlation coefficient 0.432. This comparison needs no explanation to indicate the

dominant role of the wintering conditions for the yield. But the matter is not confined merely to this.

4. We noted in our survey of spring wheat that under the conditions of the Poltava station insect pests, principally the Hessian fly, and fungus diseases -- especially rust, which is peculiar to a more humid climate -- did no slight damage to the yield of spring wheat. The same is observed with winter wheat, as may be gathered from the following table:

YIELD IN TSENTNER PER HECTARE

1 year with very serious damage (1 ball)	9.0	12.3
4 years with serious damage (2 balls)	13.2	
2 years with appreciable damage (3 balls)	19.8	
13 years with no damage (4 balls)	21.3	

The difference between the yield in the five years with serious and very serious damage, and the 13 years that were spared from pest damage (insects and fungi) reaches the level of 9 tsentner per hectare. The correlation coefficient between yield and the numerical expression of the degree of insect and fungus damage is 0.627. The combined influence of wintering conditions and insect and fungus damage on the yield is characterized by a correlation coefficient of the order of 0.627. We note one detail of no little importance: an appreciable relation is observed between the degree of winter damage and the degree of insect and fungus damage. The correlation coefficient between these series is + 0.501. An interaction apparently exists here: plants weakened by a bad wintering become less resistant to the attacks of pests, while on the other hand the pest-damage done to winter wheat in autumn lowers its hardiness against unfavorable wintering conditions; for instance, in 1902-03, the winter wheat crop,

infested in autumn by the Hessian fly, proved non-winter-hardy in the unfavorable winter of 1902-03.

Finally, we note still another phenomena which is harmful to the yield: the lodging (POLEGANIYE) of winter wheat, which in two years (1883 and 1894) (the lodging of crops is almost always connected with preceding weather that was in general favorable. This weather calls forth a stormy development of the vegetative organs of the strawy part of the plants, thus creating the danger of lodging actually does take place for any reason, even a considerable reduction of productivity will not lead to an excessively low harvest, for the level from which the yield was reduced was fairly high already. This is why the damage from lodging is not so striking as, for instance, that from drought) of the 20-year period, sharply reduced the yield. Inclusion of this third factor with the wintering damage and pest damage increases the correlation coefficient to 0.880.

The dominating combined influence of these three factors is thus immense. In comparison to it, the direct action of the autumn and spring-summer weather on the yield, of winter wheat, under the conditions of the Ukrainian wooded steppe, retires to the background. The years free from winter damage, from insect and fungus damage, and from lodging of winter wheat, (and there were 10 such years), all gave yields no lower than the average: from 18.9 to 30.15 tsentner per hectare, as against a 20-year average yield of 18.9 tsentner. Thus if the wintering conditions were favorable and the winter crops emerged from beneath the snow without damage, or with only insignificant attacks, if pests (insects and fungi) were absent or their activity was weak, and if there was no lodging during the grain-

forming phase, then an over-average yield could be almost certainly counted on in the Ukrainian wooded steppe -- but of course only on fields with high cultural practice approximating that of the series being investigated.

The autumn and spring weather, however, does not pass entirely without consequences for winter wheat: it still places its stamp on the yield, though it is considerably weaker than the combined influence of the three factors we have mentioned. The correlation coefficient is raised to 0.915 by including in the regression equation the rainfall index P, characterizing the positive influence of the autumn and spring rainfall and the negative influence of rainfall in the grain-forming phase, (instead of 0.880 for the three factors: wintering conditions, insect and fungus damage, and lodging). With this we conclude our survey of the results of the study for Poltava station (details are set forth in the special work-up for this station; see also Diagram 3).

DIAGRAM 3

YIELD OF SPRING WHEAT AT THE POLTAVA AGRICULTURAL EXPERIMENT STATION

Observed yield
 Yield calculated by the regression equation:

$$3^V = -13.9200 - 3.9464W + 1.3516B + 3.5466L + 0.1177P;$$

$$R = 0.928; R_1 = 0.915$$

Note: The yield for 1891 is not shown, since winterkilling of winter wheat took place in that year.

5. For the Sumy experiment station, situated on the more humid northern wooded steppe, under conditions of work that did not

YIELD IN TSENTNER PER HECTARE

permit my taking into the lodging of the crops, for the yield - owing to the lack of information on these subjects -- I was able to isolate two factors: (1) the autumn rainfall and (2) the spring-summer temperatures. Both factors show positive influence on the yield of winter wheat. The other basic elements of the spring-summer weather (rainfall and temperature) are either entirely ineffective or else their values, being small, are colored and obscured by the more powerful action on the yield by factors on which we have no information, i.e. by the wintering conditions, the insect and fungus damage, and the lodging. (We have seen from the example of the wooded-steppe station of Poltava how great a significance the inclusion of these factors has for the work.

The positive influence of the autumn rainfall is measured by a correlation coefficient of the order of $+0.376$, while the favorable influence of warmth during the early spring period in the life of the winter wheat is expressed by a correlation coefficient of the order of $+0.510$. The coefficients are negligible for the other meteorological elements (Table II).

[See Table II on following page]

The positive influence of warmth during the initial period of the spring life of winter wheat entirely corresponds to the climatic position of Sumy station, in which the proximity of Polesia has already been remarked. This prevalently favorable influence of warmth on the yield of winter wheat is indubitable at Sumy Station (details in the work on Sumy station). Apparently at Sumy station, where the moisture is adequate, elevated temperatures during the first half of the spring, by heating the soil and encouraging the

TABLE 11

CORRELATION COEFFICIENTS OF THE WINTER WHEAT YIELD AT SUMY
EXPERIMENT STATION WITH THE TEMPERATURE AND RAINFALL

Periods	Correlation coefficients of yield	
	WITH TEMPERATURE	WITH RAINFALL
1st period: the four decades after the onset of spring	0.540	-0.043
2nd period: from the 5th de- cade after the onset spring to the heading to	-0.100	-0.079
3rd period: from heading to harvesting	0.006	-0.096

release of excess moisture from it, create conditions more favorable to the aeration of the soil, and consequently also more favorable for soil and bacterial processes. In many cases the direct influence of low temperatures, which retard plant development and not infrequently damage the sowings, are also of decisive importance.

The Saratov experiment station, situated in the arid zone along the Volga, stands apart from all the others. In contrast to what we have observed in the Azov-Black Sea Kray and in the Ukraine, the spring-summer weather exerts an overwhelming influence on the fluctuations in the yield of winter wheat (just as it does for spring wheat) in the arid Povolzh'ye: the correlation coefficient of the May-June rainfall with the yield reaches the value of + 0.826, while for the number of days with harmful temperatures it is -0.840. The combined action of these two factors is expressed by the correlation coefficient 0.92 (materials on the Saratov station). The mean pro-

ductivity of winter wheat, however, was not low: the harvest over 15 years was 17.7 tsentner per hectare, ranging from a low of 7.6 to a high of 28 tsentner (if we except 1921, when winterkilling of winter wheat took place). Just as for spring wheat, snow retention, which strengthens the soil moisture stocks by the spring, and for good measure assures the best snow cover for the winter sowings, proved to be most effective agricultural measure, increasing the mean productivity of winter wheat to 22.2 tsentner per hectare (with a range from 10.5 to 32.8 tsentner per hectare) and thus bringing it almost up to the level of productivity for winter wheat at Poltava experiment station in a fertilized eight-field crop rotation. (Snow retention likewise increased the yield of spring wheat at the Saratov station almost to the level of productivity for spring wheat at the Poltava station in a fertilized crop rotation), (21.2 tsentner per hectare). The yield was increased by 4.5 tsentner per hectare over that obtained without the use of snow retention (22.2 tsentner against 17.7 tsentner). But while snow retention increases productivity in the arid zone, it does not completely solve the question of liquidating the drought, which both limits productivity and makes the yield unreliable. Just as for spring wheat, it is irrigation that introduces this revolution into the agriculture of the arid zone. In spite of poorer climatic and soil conditions at the Kostych experiment station, as compared to Saratov, in spite of the crude system of agriculture in use at Kostych, the mean 12-year productivity of irrigated plots there was 25.6 tsentner per hectare. It is, however, more correct to compare the yields for years in which there are yield data on both irrigated and snow-retention plots; and the 7 years from 1924 to 1930 show 26.6 tsentner per hectare

for irrigated farming at Kostych station, and 19.6 tsentner per hectare for snow-retain at Saratov station, or a difference of 7 tsentner in favor of irrigated farming. Moreover, as the study of V. Shneydyuller has shown, the yield from the irrigated plots of the Kostych station are, in the first place, completely emancipated from the influence of the rainfall (the correlation coefficient between yield and May-June rainfall being + 0.03, while it is + 0.826 at the Saratov station); and, in the second place, considerably less subject to the harmful influence of high temperatures (the correlation coefficient being -0.445 at Kostych station and -0.64 at Saratov station) and of dry winds (with correlation coefficients of -0.16 and -0.78 for Kostych and Saratov stations respectively).

This indirect influence of irrigation is explained by its beneficial influence of the microclimate of the irrigated tracts. In a word, the picture is the same for spring and for winter wheat.

7. We conclude our survey with a resume of the results at Shatilovsk station, which is situated in the wooded steppe in the North of Orel oblast, not far from the boundary of the non-chernozem zone. As far as the short 9-year series permits, features common to both Ukrainian wooded steppe and the arid southeast may be noted. The negative influence of high temperatures during the grain-forming phase, the positive influence of rainfall during this period and that preceding the heading, are features that make the Shatilovsk station similar to that at Saratov; but the harmful action of high temperatures and the favorable action of rainfall are manifested in considerably attenuated form by comparison with the Southeast. The absence of any definite, appreciable influence

of the temperature from the onset of spring to the heading, and the positive influence of autumn rainfall, are both features which recall the Ukrainian wooded steppe (but more Poltava than Sumy). The rather appreciable negative influence of early spring rainfall is peculiar to the Shatilovsk station. It is possible that in this the influence of the non-chernozem zone is already felt. I repeat, however, that these inferences are all based on an analysis of a very short series of years (nine) Details are contained in the author's unpublished study on the Shatilovsk station.

From a manuscript completed in 1937.

A CONTRIBUTION TO THE QUESTION OF THE DETERMINATION
OF THE EXPECTED YIELD

(From materials on Moscow and Petrograd Guberniyas)

I

The present article represents a short exposition, first of the author's work on the quantitative estimation of the yield to be expected according to the subjective evaluation of crop condition by the correspondents of the current agricultural statistics, and, secondly, of an experiment in determining the yield for Petrograd guberniya by methods of mathematical statistics, on the basis of the dependence between yield and meteorological factors. The brevity and schematic nature of the exposition are due to the need for compressing as great a content as possible into the small framework of the present paper.

The estimation of the expected yield from the subjective evaluation of crop conditions at blooming time was practiced by the former Central Statistical Committee from 1904 on. The following arbitrary scale was used for converting the characterizations of the correspondents into quantitative indices: "excellent" condition of the cereal crops was taken as equivalent to double the average yield during the preceding 10 years. "Very good" condition was taken as $1 \frac{2}{3}$ of that average yield; "good" was taken as $1 \frac{1}{3}$ that yield; "average" was taken to equal it; "under average" as $\frac{2}{3}$ of it; "poor" as $\frac{1}{3}$ of it; and crop failure was taken as zero.

From 1911 on, the Moscow Statistical Bureau employed an improved method of estimating the yield. The subjective characterizations of crop condition "excellent", "good", "average", "below average", and "poor", received a quantitative expression based on the empirical material. From the specific entries of crops with uniform characterizations (for instance, crop good), the mean value of the yield was derived, which then served as a quantitative measure for the corresponding subjective characterization. This system has been adopted at the present time by the Central Statistical Administration (This was written in 1921. Editor's Note).

In my work on the Petrograd guberniya, I also used the Moscow method (The choice of guberniyas (Petrograd and Moscow) for the present exploratory work depended exclusively on the fact that the author, because of his residence, was able to collect this material originally, on his own private initiative, only for these guberniyas).

The material collected by me on the condition of the rye crop in the Moscow guberniya covers a period of 20 years and refers to three stages in the development of the plant: (1) the moment of onset of winter (in what condition the winter crops went under the snow; (2) the period of early spring (in what condition the winter crops emerged from under the snow); and (3) the condition of the rye crop in the middle of June. For Petrograd guberniya the winter-crop condition data covers the 14 years from 1902 to 1915 and relates only the moment of emergence from the snow (The data on the mid-June condition of the winter crops in the Petrograd guberniya is fragmentary and disconnected. Their autumn condition, before the

onset of winter, could have been extracted from the "Collections of current statistics", but for certain years gross errors had crept into the computations, in view of which I deemed it more advisable to abstain entirely from handling this period).

Conversions of the qualitative evaluations to quantitative form, for the Moscow guberniya, are available for 10 years, and for Petrograd guberniya for 11 years, for which purpose it proved necessary to reprocess the file material of the current statistics of the Petrograd Bureau.

Let us see what the quantitative expressions of "good", "average" and "poor" yield are in both these guberniyas.

CHARACTERIZATIONS OF THE CROP	CROP "ITSELF"		"AVERAGE" YIELD = 1	
	MOSCOW GUBERNIYA	PETROGRAD GUBERNIYA	MOSCOW GUBERNIYA	PETROGRAD GUBERNIYA
Good	5.61	7.13	1.31	1.31
Average	4.15	5.46	1.00	1.00
Poor	2.80	3.46	0.65	0.63

Note: In view of the fact that the condition of the winter crops in Moscow guberniya is not indicated for all years by the characterizations "excellent", "good", "average", "below average", "poor", all characteristics had to be combined into three groups.

For the years for which the qualitative evaluation could be converted to a quantitative one, the actual average crop "itself" was 3.8 for Moscow guberniya and 5.1 for Petrograd guberniya. The estimate of the average crop by the correspondents thus approached

the actual average crop very closely, and was somewhat above it. A second inference that follows from these data is no less important: the relation of the characterizations in both guberniyas may be considered identical: a "good" crop is approximately $1/3$ higher than the average, while a "poor" crop is about $1/3$ below it (If this is confirmed by the work-up of the material for the other guberniyas, then the scheme of the Central Statistical Committee is not as arbitrary as it may seem at first sight. Finally, a further work-up of the material leads us to a third conclusion: the crop estimate fluctuates somewhat according to the years. In good years it is higher, in poor years it is lower. Apparently the estimate by correspondents of the yield on their own farms "bifurcate" and adjust themselves to two different "levels". On the one hand, the yield is compared with the "average" yield through time, while on the other hand -- perhaps involuntarily -- the yield is estimated "spatially", i.e. in relation to that of the neighboring farms. If we designate the yield by the symbol y , and the quantitative characterization of the average yield by years, by the symbol v , then the correlation coefficients between v and y are:

$$\text{For Moscow guberniya, } r_{v,y} = +0.75$$

$$\text{For Petrograd guberniya } r_{v,y} = +0.80$$

The corresponding regression equations are as follows:

$$\text{For Moscow guberniya: } v = 3.546 + 0.161 y$$

For Petrograd guberniya (the values v and y are expressed in puds per desyatina and represent the deviations from the normal productivity, which normally is determined separately for each year, in view of the uptrend of productivity there).

This task set for the present investigation is the determination of the degree of accuracy in the crop forecasts based on crop conditions during various periods of plant growth. Obviously, as the beginning of harvesting approaches, the coincidence of calculated and actual yield should increase. The accuracy of the forecast that results from the early spring condition of the winter crops should be greater than that of the estimate of expected yield made from the condition of the sowings at the onset of winter, and, in turn, should yield in accuracy to the assumed yield as determined from the condition of the rye in mid-June. The coefficient of correlation between crop condition and actually realized harvest may serve as a measure of parallelism and of the degree of coincidence. Let y be the actual yield, and u , z , and x the respective conditions of the rye crop at the onset of winter, early spring after emergency from the snow, and mid-June. We determine the crop condition by the difference between the percentage of reports of good condition of the winter crops and the percentage of reports of their poor condition, placing a plus sign when the former prevail and a minus sign when the latter prevail (Thus, for example, in 1909 the condition of winter rye on 15 June was characterized by 20.2% of good condition reports, 48.6% of average condition reports, and 31.2% of poor condition reports; and the value of x for 1909 was $20.2 - 31.2 = -9.0$. For 1915 the corresponding percentages were 48.7% good, 46.3% average, 5.0% poor; and the value of x for 1915 was $48.7 - 5.0 = +43.7$).

for condition of winter crops at the onset of winter: $r_{y,u} = +0.06$
 for condition of winter crops after emergence from snow: $r_{y,z} = +0.61$
 for condition of winter crops in mid-June: $r_{y,x} = +0.57$

The correlation coefficient between the yield and the correspondents' evaluations of the sowings at the onset of winter is so negligible that it is hard to speak of any relation at all between the values of y and u . This may be explained either by the actual failure of the rye yield to reflect the condition of the sowings towards the end of the autumn period, or, in the alternative, by the insufficient accuracy of the correspondents' estimates of the condition of the rye at the onset of winter. The former assumption would be equivalent to admitting that the conditions of rye seeding and of the autumn life of winter rye are devoid of significance for the yield of rye in Moscow guberniya. This hypothesis -- in itself improbable -- is refuted by the fact of the influence on the rye crop of the quality of the seed harvested during the previous year (not to mention the other factors of the sowing and autumnal vegetation of the winter crop). Let y be the yield, and w the weight of a measure of grain of the preceding year's harvest. Then the coefficient of correlation (This correlation coefficient is calculated from a comparison of the yield according to the data of the Central Statistical Committee with the weight of a measure of rye harvested in the preceding year, according to the Zensivo statistics. The period covered by the investigation is 33 years, consequently the standard error of the correlation coefficient is

$$\frac{1 - 0.42^2}{\sqrt{33}} = \frac{0.8236}{5.745} = 0.143,$$

i.e., the coefficient of correlation is three times its standard error. By the probability tables this means that there are only three chances (approximately) against 997 that the correlation coefficient is of accidental origin) $r_{y,w} = 0.42$. The quality of the grain, as the seeding material, affects the first stages of plant development, its germination and growth in the autumn period

the life of rye. Thus the lack of connection between the yield and the correspondents' appraisal of the autumn condition of winter rye must be attributed to the imprecision of such appraisal.

We see further that the correlation coefficients for the condition of winter rye at the onset of spring and in mid-June are almost identical (+ 0.61 and + 0.57), which does not speak in favor of the accuracy of the correspondents' appraisals for the latter period, for, as has been pointed out above, the correlation coefficient between crop condition and yield should increase as the harvest approaches. The question arises whether the relatively small value of the correlation coefficient corresponding to the last period does not depend on the method of determining the condition of the cereal crops by the differences in the percentage of good and poor condition-reports. We use the method of evaluating their condition on the basis of the qualitative characterization: "good", "average", "poor", equal for Moscow guberniya to the "self" figures of 5.61, 4.15, and 2.8 respectively. For this purpose we multiply each characterization by the corresponding percentage of reports, add the products, and divide the total by 100 (It could have been for (It could have been foreseen in advance that the quantitative characterization of a "good" yield would deviate to almost the same extent from the "average" as does the characterization of a "poor" yield. Therefore the difference between the percentage of good and poor condition reports is almost the same kind of a measure of crop condition as their estimation according to the expected yield). The indices of cereal crop condition, expressed ~~now~~ in the form of the expected yield "in itself", are given in column 6 of the table which follows:

TABLE 1 (p. 248)

DEPENDENCE BETWEEN THE YIELD ("IN ITSELF": SAM SKOL'KO) AND THE CONDITION OF THE RYE CROP DURING THE FIRST HALF OF JUNE, OLD STYLE
 y : yield "in itself" according to data of Zemstvo statistical service.

x_1 : condition of rye crop in first half of June; excess in percentage of good over poor crop condition reports denoted by + sign, the reverse denoted by - sign.

x_2 : condition of rye crop in first half of June, expressed in quantitative estimates derived from the qualitative estimates.

x_3 : the same, with correction for crop of prior year.

x_4 : the same, with correction for mean yield.

ELEMENTS FOR CALCULATION OF CORRELATION COEFFICIENTS (p. 249)

ARITHMETIC MEANS:

$$\begin{aligned} \bar{y} &= 3,9 \quad (3,88) \\ \bar{x}_1 &= + 16 \quad (16,1) \\ \bar{x}_2 &= 4,46 \\ \bar{x}_3 &= 4,36 \\ \bar{x}_4 &= 4,05 \end{aligned}$$

STANDARD DEVIATIONS:

$$\begin{aligned} \sigma_y &= 0,77981 \\ \sigma_{x_1} &= 29,365 \\ \sigma_{x_2} &= 0,409049 \\ \sigma_{x_3} &= 0,43655 \\ \sigma_{x_4} &= 0,41593 \end{aligned}$$

MEAN PRODUCTS OF DEVIATIONS:

$$\frac{\sum (y - \bar{y})(x_1 - \bar{x}_1)}{20} = + 13,312$$

$$\frac{\sum (y - \bar{y})(x_2 - \bar{x}_2)}{20} = + 0,18591$$

$$\frac{\sum (y - \bar{y})(x_3 - \bar{x}_3)}{20} = + 0,20942$$

$$\frac{\sum (y - \bar{y})(x_4 - \bar{x}_4)}{20} = + 0,20245$$

CORRELATION COEFFICIENTS:

$$r_{y, x_1} = + 0,5733$$

$$r_{y, x_2} = + 0,5858$$

$$r_{y, x_3} = + 0,6151$$

$$r_{y, x_4} = + 0,6240$$

Let us now compare the correlation coefficients $r_{y,x1}$, $r_{y,x3}$ with the correlation coefficient $r_{y,x1}$ $r_{y,x3}$ $r_{y,x4}$ where the indices $x1$, $x2$, $x3$ are the indices of condition of the winter sowings after emergency from the snow, calculated by the corresponding method.

METHOD OF DETERMINING CROP CONDITION	CROP CONDITION AT EMERGENCY FROM SNOW	WINTER CROP CONDITION AT MID-JUNE
Difference in percentage of good and poor condition	$r_{y,x1} = +0.602$	$r_{y,x1} = +0.573$
Expected yield corrected for yield of prior year	$r_{y,x3} = +0.632$	$r_{y,x3} = +0.615$
Expected yield corrected for yield of prior year and for mean yield	$r_{y,x4} = +0.637$	$r_{y,x4} = +0.624$

Thus the introduction of corrections somewhat "improves" the dependence for each period, but the basic fact still remains in force: (the correlation coefficient r that determines the parallelism between the yields and crop condition, are almost the same for the early period and for the later period (and are even somewhat higher for the earlier period). We conclude from this that the evaluation of the condition of the winter sowings in mid-June by the correspondents suffers from greater defects than their characterization after emergency from the snow (I have worked out a few other variants still confirm the basic fact established above. It is not superfluous to remark in this connection that the indices of the expected crop, calculated by the five-ball system ("excellent" yield, good, average, poor) for those years

in which the correspondents reports are so distributed, are almost no different from the indices calculated by the three-ball system (good, average, poor). Thus in any case the relative small value of the correlation coefficients cannot be explained by the summary nature of the method used to determine the expected yield).

Up to now we have been comparing the yield of rye according to the date of the ^{ed}zenstvo statistics, express/in the form "in it-self", with the correspondents assessments of the condition of the rye at different period of development. Let us now find the correlation coefficients between the yield of rye in puds per deayatin according to the date of the Central Statistical Committee, with the same indices z_4 and x_4 and compare them with the preceding coefficients.

PERIOD TO WHICH THE ESTIMATES APPLY	CORRELATION COEFFICIENTS FOR YIELD, ACCORDING TO DATA OF TASK		CORRELATION COEFFICIENTS FOR YIELD ACCORDING TO DATA OF ZENSTVO STATISTICAL SERVICE
Emergence from the snow	$r_{y,z_4} = + 0.649$		+ 0.637
Mid-June	$r_{y,z_4} = + 0.654$		+ 0.624

It will be seen from this table that the correlations coefficients are almost identical. The new variant thus confirms the former conclusions.

But could it perhaps be the case that each of the indices z_4 and x_4 catches peculiarities that escape from the others? If this suggestion is true, then the combined determination of the expected yield from the values of z_4 and x_4 should improve the parallelism

between the estimate of the condition of rye and the actual yield. For this purpose we find the regression equation that determines the expected yields in dependence on the condition of the winter rye as estimated by the correspondents (For this purpose we find the auxiliary coefficient of correlation $r_{x_1, y_1} = 0.826$, and then use the formula given in the noted textbook of Slutskiy. The same regression equation may also be obtained by the method of least squares.

$$y_1 = 46.4 + 8.48 (x_1 - 4.05) + 6.27 (x_1 - 3.83)^2$$

In the following table the values of the expected yields y_1 are compared with the actual yield, y (alpha being the difference).

Year	y	y_1	$\alpha = y - y_1$	Year	y	y_1	$\alpha = y - y_1$
1	2	3	4	1	2	3	4
1897	40.6	47.0	-6.4	1907	30.5	34.8	-4.3
1898	49.1	46.3	+2.8	1908	35.7	33.1	+2.6
1899	55.2	47.9	+7.3	1909	23.9	35.1	-11.2
1900	54.6	50.6	+4.0	1910	48.3	48.4	-0.1
1901	48.4	49.3	-0.9	1911	53.9	53.8	+0.1
1902	44.9	44.2	+0.7	1912	57.1	56.5	+0.6
1903	44.4	38.6	+5.8	1913	52.0	49.3	+2.7
1904	58.6	51.2	+7.4	1914	42.7	44.3	-1.6
1905	52.9	42.6	+10.3	1915	58.0	51.4	+6.6
1906	32.6	52.8	-20.2	1916	45.6	51.8	-6.2

MEAN $y_0 = 46.4$
Standard deviation $\sigma_y = 9.5431$
Mean deviation of y series:

$$\sigma = 7.725$$

$$\sigma_\alpha = \sqrt{\frac{\sum \alpha^2}{20}} = 6.96$$

Mean deviation of absolute values of alpha

$$\beta = \frac{\sum |\alpha|}{20} = 5.09$$

The values of alpha denote the deviations of the actual productivity from the expected productivity, calculated from the regression equation. The standard deviation of the series alpha is $\sigma_{\alpha} = 6.76$ (The value of σ_{α} may be calculated either by the Slutsky formula or by the formula of the method of least squares (from the coefficient of the regression equation), while the mean deviation, denoted by the symbol β , and defined as the arithmetic mean of the absolute values of alpha ($\beta = \frac{\sum |\alpha|}{20}$) is equal to 5.09. Consequently the mean error, made by us in determining the yield from the estimates of crop condition in early spring and at mid-June, is 5.09 puds.

Let us see how much the determination of the expected yield by the regression equation represented above has been "improved". For this we find the correlation coefficient between the values of y (that is, the indices of the expected yield) and the series y_1 (that is, the indices of actual yields): $r_{y,y_1} = +0.682$ (The value of r_{y,y_1} may be calculated either by the usual method or (what is far simpler) by the formula $r_{y,y_1} = \sqrt{\frac{\sigma_y^2}{\sigma_{y_1}^2}}$).

As we see, the "improvement is very insignificant, for the values $r_{y_1,sh}$ and $r_{y_1,hd}$ are only slightly smaller than r_{y_1,y_1} (+ 0.649 and + 0.659).

Both in the period of emergency of the winter rye from beneath the snow, and also to a considerable extent in mid-June, the correspondents must judge the condition of the crop from the external appearance of the "straw" part of the plant. The "mystery" of grain formation still yields no signs to the external senses. Moreover, the quantitative relation between the harvest of grain and

the harvest of straw change violently from year to year. According to the 27-year data of the Central Statistical Committee, from 1889 to 1915, 61 puds of grain are harvested on the average for every 100 puds of straw, but this "relative yield of grain fluctuates sharply from year to year within a range of 43 to 75 puds of rye grain to 100 puds of straw. Consequently the observations of the correspondents at these moments in the life of winter rye are necessarily one-sided and incomplete, in consequence of which fact errors are unavoidable in determining the expected yield of grain on the basis of crop-condition reports (The observations of correspondents at mid-June (Old Style) do not always cover even most of the blooming period of the rye. Although rye begins to bloom on 4th of June (Old Style) in Moscow guberniya, and ends on 18 June, the annual fluctuations in the blooming period are very considerable). In other years the blooming of rye is very late (beginning 17-18 June), and in still others it blooms very early (16 May) thus the observations of the correspondence, which are set for 15 June, do not in general cover the middle of the ripening period, when it already becomes possible to form an opinion from external impressions directly (and not indirectly) on the expected yield of grain. I do not consider it superfluous to remark that the spring weather, long before blooming, already includes elements of the relative yield (i.e., the yield of grain to 100 puds of straw), but of course this is revealed only in the period of grain formation. At least I succeeded of noting this fact for the Petrograd guberniya where the May-June rainfall is even more strongly reflected in the relative yield of grain than the July rainfall (the dates of rainfall are always given in New Style), when the ripening and maturation of the grain takes place. If we denote the yield of rye grain per 100 puds of straw by the symbol $\frac{g}{x}$ and

the rainfall of May, June and July (New Style) by the symbols

respectively, I found that

$$r_{PYI} = -0.60$$

$$r_{PYI} = -0.64$$

$$r_{PYII} = -0.36$$

the influence of rainfall on the relative yield of grain is negative for all months, but is weaker in July than in May or June.

in all probability soil processes play a role here.

If this hypothesis is true, then there should be parallelism between the values of (indices of the spread between expected and actual yield, cf. Table 2) and the relative grain yield (i.e. the yield of grain to 100 puds of straw). The errors in the determination of the expected yield should be the larger, the more the relative yield of grain (not given by the correspondents' observations during the periods under study) deviates from the normal. In years of good relative grain yields, a tendency to underestimation of the yield should of course be noted, while in years of poor relative yields the calculated indices of expected yield should, in general, give a picture of too optimistic estimates. This parallelism exists in fact: the correlation coefficient between and 2 q (relative yield) is:

$$r_{q, y} = +0.511.$$

The corresponding regression equation is:

$$\hat{y}_1 = +0.24 + 0.404 (q - 58).$$

We correct the indices of expected yield (series y_1 in Table 2)

by using this equation of regression. As a result we obtain the new series y_2 , giving the expected yield under the assumption that the annual fluctuations in the relative grain yield to 100 puds of straw had been registered by the correspondents (see Table 3).

As may be seen from the data of this table, these corrections to the indices of expected yield lead to a greater coincidence of expected and actual yields: the coefficient of correlation is increased to from 0.68 to 0.77. The standard deviation is reduced from 6.97 to 5.94 puds, while the mean deviation declines from 5.09 to 4.38 puds. Thus the non-coincidence of the expected yield, determined by the correspondents' estimates of crop condition in the period of emergence from the snow and in mid-June, is partially explained by the fact that at these moments of plant life the correspondents can judge the expected grain yield only by the condition of the "straw" part of the plant, for the grain itself has not yet been formed. But we cannot say that this divergence of the expected and actual yields is entirely or even for the greatest part explained by this circumstance (If the correlation coefficient $r_{x, y}$ were near unity, which would inevitably involve insignificant values of β in column 4, we would theoretically have the right to say that the divergence between the expected and actual yields had been entirely, or almost entirely, explained by the impossibility of the correspondents' judging the relative grain yields at the times in question. If $r_{x, y} > 0.71$ we could maintain that the non-coincidence depended for the largest part on this circumstance. In our case $r_{x, y} = 0.51$. The value of this coefficient allows us to say that only a quarter of the error

in the determination of the expected yield is accounted for by the early date of the observations, which do not permit judging of the relative yield of grain.) The non-coincidence between the series y_2 and the series y still remains too great for that (see series β in Table 3)

Year	$Y_2 = y_1 + d$	y	$y - y_2$	
1897	54,2	40,6	-13,6	y_1 : expected yield
1898	51,1	49,1	-2,0	d : correction
1899	51,5	55,2	+3,7	y_2 : expected yield after correction
1900	53,8	54,6	+0,8	y : actual yield
1901	51,3	48,4	-2,9	: divergence between actual and expected yields
1902	43,7	44,9	+1,2	Standard deviation of the values of β
1903	38,5	44,4	+5,9	$\sigma_\beta = 5.943$
1904	54,8	58,6	+3,8	Mean deviation of the absolute values of β
1905	46,6	52,9	+6,3	$\bar{\beta} = \frac{\sum \beta }{19} = 4.38$
1906	47,9	32,6	-15,3	Correlation coefficient between y and y_2
1907	29,1	30,5	+1,4	$r_{y,y_2} = 0.77$
1908	28,2	35,7	+7,5	
1909	30,2	23,9	-6,3	
1910	46,7	48,3	+1,6	
1911	53,3	53,9	+0,6	
1912	54,8	57,1	+2,3	
1913	50,4	52,0	+1,6	
1914	43,8	42,7	-1,1	
1915	53,0	58,0	+5,0	

The question arises whether the considerable divergency of the expected and actual yields is not to be explained by the small number of correspondents' observations. Their number does in fact

fluctuate greatly over the 20-year period under study between 112 and 1297 for early spring, between 103 and 1004 for mid-June). Consequently the element of chance could play an appreciable role in the determination of the expected yield and could introduce appreciable errors in its indices. Let us test this hypothesis. If it is true, then the years with fewest observations should show the greatest divergences between expected and actual yields. While the years with the most correspondents' reports should be characterized by the lowest errors. In the former case the values of series beta (table 3) taken without regard to sign, should be greater than in the latter. We find the coefficient of correlation between the absolute values of beta (i.e. regardless of plus and minus signs) and the square root of the number of observations. An inverse character for the dependence of the errors in the expected yield on the number of observations should be marked by a minus sign before the correlation sign, which is, in fact, observed. But the value of the correlation coefficient is very small:

$$r_{\beta\sqrt{n}} = -0.26,$$

where n is the number of observations. The size of this correlation coefficient lies completely within the range of its random variation. Consequently we are entitled to draw the following conclusion: the errors in the expected yield, established on the basis of crop-condition reports from correspondents at the moment of emergence from the snow and in mid-June, can be explained to a very small extent by the insufficient number of observations.

Obviously collective errors in crop evaluation must primarily insinuate themselves into the correspondent reports.

It would be a very interesting task to discover the "regularities" of these collective errors and determine their origin, but a separate investigation would be necessary to solve this problem, which study would have to be based on more extensive material than that of which the author disposes at present. However, even from the available data, certain conclusions, which of course are of a hypothetical, conjectural nature. Thus, for example, it was noted above that the size of the yield is reflected in the correspondents' quantitative estimates of it. It may be expected by analogy that the varying crop condition is not without influence on the estimates of it, and namely: when crop conditions are good they are in general evaluated more strictly, while when they are bad the characterizations are more "lenient". Moreover, it is probable that the preceding stage of plant development is reflected in the evaluation of crop condition for the subsequent period: improvement in crops after a poor condition is viewed more optimistically by correspondents that it should be, and, conversely, a deterioration following a good crop condition is outlined in a darker light than the facts warrant. Finally, it is very possible that the impression of preconceived ideas on the meteorological elements that influence plant development may be stamped on the correspondent estimates. There are, however, certain specific conditions that influence the degree of reliability of the crop-condition reports in the various periods of plant life. As has already been pointed out, the correspondent estimates for mid-June are less satisfactory than those for the moment of emergency from the snow (The correlation coefficients are the same for these two periods, but it should be higher for mid-June, since at that time the harvest is nearer).

Let us try to explain this fact, insofar as we can.

A stationary and elementary nature of the observations corresponds to crop condition at emergency from the snow. In spring and summer, however, the correspondents must observe plants in the process of growth and development. The spring weather at emergency from the snow may have either a positive or a negative influence, but even in the latter case the winter crops continue to grow and develop, and in this sense their condition "improves". As a result a tendency to optimistic estimates develops, since it is difficult on the basis of external impressions to determine the degree of normality in this process of "growth -- improvement", or to say whether it appears to be below normal, normal or above normal. And the correspondent estimates for mid-June are in fact higher than the quantitative characteristics of crop condition at emergency from the snow.

It will be seen from Table 4 that in 16 cases out of 20 the later estimates of crop condition are higher than those at emergency from the snow. The mean expected yield over 20 years, determined from the June 15 crop condition, is expressed by the "crop itself" index of 4.05, while for the early period this index is 3.83 (In all probability one of the reasons for the unsatisfactory nature of the crop condition reports at the onset of winter is the same circumstance: in the autumn period the correspondents must observe the winter crops in the process of development. Even under the most unsatisfactory weather conditions, they still do grow and develop, and this process of growth of the winter crops becomes fixed in the consciousness of the correspondents as an improvement in condition. A strong tendency to optimistic evaluation results.

The difference between the percentage of reports of good crop condition and that of reports of poor crop condition is 30% over the 20-year period). (see Table 4).

The inferior quality of the data relating to mid-June apparently also depends on the circumstance that at this period the correspondents have no sufficiently elementary objective criterion, readily perceptible to the external senses, for characterizing the expected grain yield. There are, however, such criteria in the early period for judging the condition of the winter crops: for it is easy enough to determine ^{from the} half-covered and bare spots formed, how much the crops have suffered during the wintering period; and in just the same way it is easy to establish the condition of the undamaged parts of the winter crop from their height and bushiness. By mid-June, when the winter crops have already eared, and are in most cases in the blooming stage, the correspondents must combine a number of criteria to judge the probable grain yield: note the spacing and height of plants, their spike density, the quality of the spikes, determine how the blooming season has been, and then as a result prepare a "diagnosis" as to what condition of the winter crops is observed in the present year, whether "good", "average" or "poor". The mid-June estimates of the correspondents are thus the consequence of a fairly complicated process of inference and are not given by direct observation.

EXPECTED YIELD			EXPECTED YIELD		
From estimate of condition of winter rye at emergence			From estimate of condition of winter rye on 15 June (Old Style)		
YEAR	from the snow		YEAR	winter rye on 15 June (Old Style)	
[1]	[2]	[3]	[1]	[2]	[3]
1897	3.88	4.08	1908	2.91	3.16
1898	3.66	4.17	1909	2.63	3.61
1899	3.84	4.22	1910	4.04	4.12
1900	4.04	4.38	1911	4.30	4.57
1901	3.94	4.31	1912	4.36	4.85
1902	3.66	3.91	1913	4.40	3.97
1903	3.22	3.58	1914	3.62	3.95
1904	4.37	4.22	1915	4.17	4.40
1905	3.56	3.80	1916	4.45	4.23
1906	4.44	4.36	Average for		
1907	3.13	3.19	20 years	3.83	4.05

It is only natural that this "diagnosis" should bear the stamp of preconceived ideas and of all possible "levels" to which the crop condition must be compared. The inaccuracy of the estimate is deepened by the circumstance that to a certain extent the crop-condition reports at mid-June are not comparable, since they do not always relate to the same vegetative phase: although in most cases rye is at the end of the blooming period by mid-June, in some years it has only commenced blooming at this time, while in others, with early and warm spring, it is already in the ripening stage. These, then, are the conditions responsible for the relatively lower accuracy of the mid-June correspondent estimates of crop condition (I do not think that this is superfluous if I note once again that I assign only conditional, hypothetical significance to these conclusions for the time being, pending work-up of the question on the basis of more extensive material.

It follows from all that has been said that the data on rye crop condition as of 1 July, (Old Style), or 15 July (New Style), should be distinguished by far greater accuracy than for the preceding period, since by this time the rye is already in the ripening stage, and consequently the correspondents have an objective criterion for judging the expected grain yield (This criterion, the grain being formed, is important not only because it makes it possible to judge directly of the expected yield of grain, but also because by its concrete and graphic nature it restrains within certain limits the influence of the various "levels" and preconceived ideas on the crop estimates. Since 1917, semi-monthly crop-condition data have been collected on an all-Russian basis. Crop forecasting from subjective reports of correspondents is thus in a

better position now than in former times. There can be no doubt that it would be even more improved if the data were related not to calendar dates but to vegetative periods. Thus, for instance, it would be useful to introduce a blank for reporting crop condition with the direction that it was to be filled out as of a week after the end of blooming.

With this we conclude our study of Moscow guberniya and pass now to Petrograd guberniya.

As has been indicated above, there are data on the condition of the winter crops for 14 years in Petrograd guberniya, which are limited to the period of the emergence from the snow. The following table groups the various methods of evaluating crop condition and gives the correlation coefficients between actual yield and the indices of crop condition.

[see next page for Table 5]

The first two columns give the yield of rye according to the data of the Central Statistical Committee, and the deviation (δ_y) of the yield from the 14-year mean. Columns 4 and 5 give a characterization of winter-crop condition by the difference between the percentage of good and poor crop-condition reports. In columns 6 and 7 the estimate of winter rye is expressed in the form of the expected rye yield, as was done for Moscow guberniya, with the single difference that here the mean yield is given in puds per desyatin (49.5 puds for the 14 years) while by multiplying the mean yield by 0.63 and 1.31 respectively. Columns 8 and 9 include the indices of expected yield, adjusted for the uptrend of productivity in

YEAR	y	Sy	X₁	SX₁	X₂	SX₂	X₃	SX₃	X₄	SX₄	L	Si	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]
1905	32.7	-16.8	-46	-59	40.6	-9.2	40.3	-11.35	38.7	-12.5	71	48	
1903	39.6	-9.9	-35	-48	42.6	-7.9	41.5	-10.15	39.6	-11.6	43	+20	
1902	41.5	-8.0	-46	-59	40.6	-9.9	39.3	-12.35	41.7	-9.5	26	+3	
1904	42.9	-6.6	-48	+35	56.5	+6.0	55.6	+3.95	52.1	+0.9	6	-17	
1906	44.7	-4.8	+67	+54	59.5	+9.0	59.5	+7.85	52.4	+1.2	8	-15	
1907	45.6	-3.9	-14	-27	46.1	-4.4	46.5	-5.15	44.9	-6.3	26	+3	
1909	48.6	-0.9	-42	-55	41.3	-9.2	42.3	-9.35	45.1	-6.1	33	+10	
1913	51.1	+1.6	+25	+12	52.5	+2.1	55.5	+3.85	58.1	+6.9	31	+8	
1910	52.8	+3.3	+15	+2	50.8	+0.3	52.5	+0.85	51.8	+0.6	26	+30	
1911	54.1	-4.6	+31	+18	53.6	+3.1	55.8	+4.15	56.5	+5.4	13	-10	
1914	58.0	+8.5	-44	+31	55.5	+5.0	59.0	+7.35	58.7	+7.5	13	-10	
1912	58.4	+8.9	+39	+26	55.1	-4.6	57.5	+5.85	58.8	+7.6	9	-14	
1908	59.0	+9.5	+53	-40	57.3	+6.8	58.2	+6.55	56.5	+5.3	7	-16	
1915	63.9	+14.4	-43	+30	55.6	+5.1	59.6	+7.95	61.9	+10.7	9	-14	

TABLE 5 (cont'd)

Average

Mean

Deviation

$\bar{y}_0 = 49.4$

$\sigma_{x1} = +13$

$\sigma_{x2} = 50.5$

$\sigma_{x3} = 51.65$

$\sigma_{x4} = 51.2$

$i_0 = 23$

$\sigma_y = 8.496$

$\sigma_{x1} = 39.37$

$\sigma_{x2} = 6.611$

$\sigma_{x3} = 7.572$

$\sigma_{x4} = 7.531$

$\sigma_i = 17.438$

[1] Sums of the Products of Deviations

$\sum \delta y \delta x_1 = +3039.5$

$\sum \delta y \delta x_2 = +500.16$

$\sum \delta y \delta x_3 = +674.01$

$\sum \delta y \delta x_4 = +798.88$

$\sum \delta y \delta i = -1540.4$

$\sum \delta i \delta x_5 = -1458.4$

$\sum \delta i \delta x_4 = -1386.7$

[2] Correlation Coefficients

$r_{y, x_1} = +0.649$

$r_{y, x_2} = +0.636$

$r_{y, x_3} = +0.748$

$r_{y, x_4} = +0.892$

$r_{x_1, i} = -0.700$

$r_{x_3, i} = -0.789$

$r_{x_4, i} = -0.754$

- ~~_____~~
~~_____~~
- y : yield in puds, according to data of Ts.S.K.
- x_1 : excess of percentage of reports of good crop condition over percentage of reports of poor crop condition (+) or the opposite phenomenon (-);
- x_2 : condition of rye at emergence from the snow, on conversion of qualitative evaluation to quantitative;
- x_3 : the same, corrected for uptrend in mean yield;
- x_4 : the same, corrected for uptrend in mean yield and adjusted for the yield of the prior year;
- i : percentage of reports that reseeding of the winter sowings with spring cereals was necessary.

Petrograd guberniya, where, in contrast to Moscow guberniya, an increase in productivity is observed. Thus in Petrograd guberniya a certain norm of productivity corresponds to each year. In Chapter 2 the methods of determining these norms for each year are set forth, and the norms themselves presented. To calculate the expected yield from the quantitative expression of the mean winter-crop condition, the normal productivity corresponding to each year is taken, and the poor or good condition of the crop is estimated by multiplying it by the multipliers just given. Since the estimate of condition fluctuates according to the yield, columns 10 and 11 give the indices of productivity with supplementary adjustments for the yield of the prior year, for which purpose the regression equation given at the beginning of this paper had to be used. Finally, columns 12 and 13 contain data on the reseeding of winter crops with spring cereals. Since these data can characterize the condition of the winter crops at their emergence from the snow, and especially that part of the winter crops that is either in very bad condition or has completely perished, I felt that their inclusion was not superfluous.

The correlation coefficients between the yield and crop condition at emergence from the snow are as follows:

Method of determining crop condition	Correlation coefficients
By percentage differences	$r_{y,x_1} = + 0.649$
Expected yield (unadjusted)	$r_{y,x_2} = + 0.635$
Same, adjusted for uptrend in yield	$r_{y,x_3} = + 0.748$
Same, with supplemental adjustment for yield of prior year	$r_{y,x_4} = + 0.892$
Percent of reseeding reports	$r_{y,i} = 0.700$

Thus the first two methods of characterizing crop condition, as should have been expected, give almost identical correlation coefficients. But adjustment for the increasing mean productivity and for the yield of the prior year powerfully "improve" the dependence of yield on crop condition at emergence from the snow; the correlation coefficients are raised to +0.748 and +0.892. Let us see how the inclusion of an additional criterion, in the form of the percent of reports of reseeded of winter crops by spring cereals, affects the accuracy of the estimates of winter-crop condition. For this we find two regression equations: one, determining the yield (y) as a function of the values of x_3 and i , and the other establishing the dependence of y on x_4 and i .

$$y_1 = 22,678 + 0,5821 x_3 - 0,1111 i$$

$$y_2 = 1,47 + 0,9522 x_4 - 0,031 i$$

The indices of expected yield calculated on the basis of these equations are presented in the following table.

Year	y	y_1	y_2	$y - y_1$	$y - y_2$
1902	41,5	41,9	40,4	-0,4	+1,1
1903	39,6	40,7	37,8	-1,1	+1,8
1904	42,9	54,2	50,9	-11,3	-8,0
1905	32,7	36,0	36,1	-3,3	-3,4
1906	44,7	56,2	51,1	-11,5	-6,4
1907	45,6	46,0	43,3	-0,4	+2,3
1908	59,0	55,6	55,0	+3,4	+4,0
1909	48,6	42,6	43,3	+6,0	+5,3
1910	52,8	49,6	50,0	+3,2	+2,8
1911	54,1	53,3	55,0	+0,8	-0,9
1912	58,4	54,9	57,2	+3,5	+1,2
1913	51,1	50,6	55,8	+0,5	-4,7
1914	58,0	55,2	57,0	+2,8	+1,0
1915	63,9	56,1	60,1	+7,8	+3,8

The correlation coefficients between yield and the respective indices of expected yield Y_1 and Y_2 are as follows:

$$r_{y, y_1} = +0.771$$

$$r_{y, y_2} = +0.893$$

Thus the inclusion of the supplemental criterion i has almost no effect on the value of the correlation coefficients, and, consequently, increases only slightly the accuracy of the crop condition estimates at emergence from the snow. Their characterization in terms of "good", "average" and "poor" proves sufficient.

The correlation coefficients that characterize the connection between the yield of rye and the crop condition at emergency from the snow are considerably higher for Petrograd guberniya than for Moscow guberniya. In all probability this may be explained by the fact that in Petrograd guberniya the conditions of the autumn weather and the winter apparently influence the rye yield more strongly than in Moscow guberniya, where the center of gravity of the influence of meteorological factors may be assumed to be relatively displaced towards the side of spring. However, only further investigation can finally answer this question. But in any case one fact may be considered positively established for Petrograd guberniya: that of the prevailing influence on the yield of rye of the ensemble of meteorological factors of the summer in the year preceding the harvest, and of the autumn and winter.

It is obvious from all that has been set forth in this chapter that all possible "levels" and collective errors are reflected in the correspondents' characterizations of crop yields, and that such characterizations are not of uniform value: insofar as the present

paper discloses, only the estimates of crop condition at emergence from the snow may be considered satisfactory, while the later reports (mid-June, Old Style) awaken more doubt with respect to their accuracy (In view of the fragmentary nature of the materials on Petrograd guberniya for this period, I have unfortunately been compelled to refrain from processing it. My general impression from glancing over it is that it confirms the conclusions drawn for Moscow guberniya). The attempt to determine expected yield on the basis of objective criteria therefore appears to be entirely in place. That basis is the march of the meteorological elements, which almost entirely determine the yearly fluctuations of productivity. An attempt at such a work-up was made for Petrograd guberniya, and gave very satisfactory results. The second chapter of this paper is devoted to the exposition of those results.

II.

This study is based on the data of the Central Statistical Committee for the entire period during which accurate crop statistics existed, i.e. from 1883 to 1916 inclusive (34 years).

Analysis of the materials by the methods of mathematical statistics discloses a steady rise in the productivity of rye in Petrograd guberniya; the correlation coefficient between the yield (y) and the time (x) is

$$r_{yx} = +0.467$$

while the corresponding regression equation is expressed by the formula:

$$y_1 = 39.92 + 0.3976x$$

The values of y_1 as calculated by this equation are shown in Column 6 of Table 1 and represent the indices of normal productivity for each year. Column 7 gives the yearly deviations of yield from the norms corresponding to that year. It is this series which we shall now subject to further investigation. We shall term it, for simplicity, the yield series or simply the yield, and shall invariably designate it by the symbol y .

[See Table 1 on the following page]

We commence our study with the rainfall during and immediately

- -

TABLE 1

Year	[1]	[2]	[3]	[4]	[5]	[6]	[7]
1883	1	- 16.5	40.4	- 6.5	40.3	+ 0.1	
1884	2	- 15.5	33.5	- 13.4	40.7	- 7.2	
1885	3	- 14.5	39.6	- 7.3	41.1	- 1.5	
1886	4	- 13.5	45.6	- 1.3	41.5	+ 4.1	
1887	5	- 12.5	49.9	+ 3.0	41.9	+ 8.0	
1888	6	- 11.5	50.7	+ 3.8	42.3	+ 8.4	
1889	7	- 10.5	45.6	- 1.3	42.7	+ 2.9	
1890	8	- 9.5	53.3	+ 6.4	43.1	+ 10.2	
1891	9	- 8.5	52.2	+ 5.3	43.5	+ 8.7	
1892	10	- 7.5	39.8	- 7.1	43.9	- 4.1	
1893	11	- 6.5	38.7	- 8.2	44.3	- 5.6	
1894	12	- 5.5	32.8	- 14.1	44.7	- 11.9	
1895	13	- 4.5	29.8	- 17.1	45.1	- 15.3	
1896	14	- 3.5	53.6	+ 6.7	45.5	+ 8.1	
1897	15	- 2.5	47.2	+ 0.3	45.9	+ 1.3	
1898	16	- 1.5	50.3	+ 3.4	46.3	+ 4.0	
1899	17	- 0.5	37.7	- 9.2	46.7	- 9.0	
1900	18	+ 0.5	55.1	+ 8.2	47.1	+ 8.0	
1901	19	+ 1.5	55.2	+ 8.3	47.5	+ 7.7	
1902	20	+ 2.5	41.5	- 5.4	47.9	- 6.4	
1903	21	+ 3.5	39.6	- 7.3	48.3	- 8.7	
1904	22	+ 4.5	42.9	- 4.0	48.7	- 5.8	
1905	23	+ 5.5	32.7	- 14.2	49.1	- 16.4	

[1]	[2]	[3]	[4]	[5]	[6]	[7]
1906	24	+ 6.5	44.7	- 2.2	49.5	- 4.8
1907	25	+ 7.5	45.6	- 1.3	49.9	- 4.3
1908	26	+ 8.5	59.0	+ 12.1	50.3	+ 8.7
1909	27	+ 9.5	48.6	+ 1.7	50.7	- 2.1
1910	28	+ 10.5	52.8	+ 5.9	51.1	+ 1.7
1911	29	+ 11.5	54.1	+ 7.2	51.5	+ 2.6
1912	30	+ 12.5	58.4	+ 11.5	51.8	+ 6.6
1913	31	+ 13.5	51.1	+ 4.2	52.2	- 1.1
1914	32	+ 14.5	58.0	+ 11.1	52.6	+ 5.4
1915	33	+ 15.5	63.9	+ 17.0	53.0	+ 10.9
1916	34	+ 16.5	50.1	+ 3.2	53.4	- 3.3

Mean $\bar{x} = 17.5$

Standard Deviation of $\bar{x} = 9.311$

Mean (yield) $\bar{y} = 46.88$

Standard Deviation of $\bar{y} = 8.363$

Regression equation: $\bar{y}_1 = 39.22 + 0.3976x$.

$$\frac{\sum \delta y \delta x}{n} = +38.24$$

$$r_{yx} = +0.467$$

Editorial Note: Column 6 is obtained from the preceding columns by calculating the straight line trend by the method of least squares. The use of the terminology of correlation theory appears logically illicit, since only in exceptional cases may this be applied to random phenomena fluctuating irregularly. Nowhere in his later works does the author admit this procedure.

preceding the sowing period. According to the samstvo statistics of Petrograd guberniya, the average date of intensified sowing there falls on 25 August New Style, while the last plowing for rye takes place during the last days of July (on the 27th to the 29th). The correlation coefficients between yield and the rainfall of several decades, beginning with July, are as follows:

1st July decade	$r_{y,1} = -0.10$	highest values of the correlation coefficients
2nd July decade	$r_{y,2} = -0.09$	
3rd July decade	$r_{y,3} = -0.33$	
1st August decade	$r_{y,4} = -0.22$	
2nd August decade	$r_{y,5} = -0.15$	
3rd August decade	$r_{y,6} = -0.52$	
1st September decade	$r_{y,7} = -0.12$	
2nd September decade	$r_{y,8} = -0.14$	

The minus sign before the correlation coefficients indicates the inverse relation between the rye yield and the amount of rainfall. In general the ten-day correlation coefficients are very insignificant and only the circumstance that all of them have the minus sign allows us to infer a negative influence of the rainfall during the period under study. The irregular character of the variation in the correlation coefficients indicates that the decade periods are too short to allow the weather in any of them to make a decided impression on the yield, and makes it necessary for us to combine the 10-day intervals into 20-day intervals (1).

(1) Even in these fluctuations, however, ascertain regularity may be discerned: the maximum values of the correlation coefficients fall within the 40-day period beginning with the third July decade

and ending with the last decade of August, i.e., the period of the final plowing for rye and that of the sowing.

1st and 2nd July decades	$r_{y, \frac{1+2}{VII}} = (-)0.16$
3rd July decade and 1st August decade	$r_{y, \frac{3}{VII} + \frac{1}{VIII}} = (-)0.40$
2nd and 3rd August decades	$r_{y, \frac{2+3}{VIII}} = (-)0.45$
1st and 2nd September decades	$r_{y, \frac{1+2}{IX}} = (-)0.13$

The dependence of the yield on the rainfall of the 20-day periods takes on a clear and entirely definite character. A strong increase in the negative effect of rainfall is observed in the period of the 3rd July and 1st August decade, from the moment that the intensified plowing for rye begins, and that immediately preceding it. The period of sowing and the decades nearest it (2nd and 3rd in August) bring a further rise in the correlation coefficient. The 1st and 2nd September decades (a period which may be termed the period of germination to the onset of tillering), for on the average the seedlings begin to emerge on 31 August to 1 September, and the rye commences to tiller on 20 September, (2) is characterized by a sharp fall in the correlation coefficient from - 0.45 to - 0.13. A particularly strong negative effect of rainfall is thus disclosed during the 40-day period covering the final plowing for rye, its sowing, and the appearance of the seedlings. The correlation coefficient between yield and the aggregate rainfall from the third July decade to the third August decade, inclusive, is - 0.54.

(2) There is no information on the vegetative periods for

Petrograd guberniya. But the data for Moscow guberniya indicate that on the average rye germinates 6 days after sowing, while tillering begins 19 days after the emergence of the seedlings. These conclusions may be boldly applied to Petrograd guberniya as well, for the time of sowing and the temperature conditions of both guberniyas are almost identical:

Guberniya	Average date of intensified sowing	August temperature	September temperature
Moscow	25 August	+15.9	+10.1
Petrograd	26 August	+15.3	+10.3

Let us apportion the rainfall more precisely among the vegetative periods. The intensified sowing of rye begins on 20 August (New Style) and ends on 31 August. The early sowings germinate on the average on 26 August, and the late sowings on 6 September. Let us denote the rainfall during the respective decades of August and September by the symbols:

$$\left(\frac{1}{VIII} \right) \left(\frac{2}{VIII} \right) \left(\frac{3}{VIII} \right) \left(\frac{1}{IX} \right) \left(\frac{2}{IX} \right) \left(\frac{1}{IX} \right)$$

Then the rainfall during the days from just before the early sowing down to germination may be taken as equal to $\frac{1}{2} \left(\left(\frac{2}{VIII} \right) + \left(\frac{3}{VIII} \right) \right)$

while the rainfall of the analogous period for the late sowings will be $\frac{1}{2} \left(\left(\frac{3}{VIII} \right) + \left(\frac{1}{IX} \right) \right)$. Consequently the rainfall for the entire period of sowing will be $\frac{1}{2} \left(\left(\frac{2}{VIII} \right) + \left(\frac{3}{VIII} \right) \right) + \frac{1}{2} \left(\left(\frac{3}{VIII} \right) + \left(\frac{1}{IX} \right) \right) =$

$\frac{1}{2} \rho_{\frac{2}{VIII}} + \rho_{\frac{3}{VIII}} + \frac{1}{2} \rho_{\frac{1}{IX}}$, i.e., it is equal to the early rainfall of the 3rd August decade plus half the sum of the rainfall of the second August and first September decades. By analogous procedure we find the formula for the period of the last plowing for rye

$$\rho_{\sim} = \frac{1}{2} \left(\rho_{\frac{3}{VII}} + \left(\rho_{\frac{1}{VIII}} + \frac{1}{2} \rho_{\frac{2}{VIII}} \right) \right)$$

and the formula for the rainfall of the period from germination to the inception of tillering

$$\rho_v = \frac{1}{2} \rho_{\frac{3}{VIII}} + \frac{3}{4} \left(\rho_{\frac{1}{IX}} + \rho_{\frac{2}{IX}} \right) + \frac{1}{4} \rho_{\frac{3}{IX}}$$

The correlation coefficients between yield and rainfall are:

Sowing period	$r_{y, \rho_c} = -0.50$
Period of final plowing	$r_{y, \rho_{\sim}} = -0.35$
Period from germination to inception of tillering	$r_{y, \rho_v} = -0.23$

Correlation coefficients constitute an exact measure of the connection between statistical series where a rectilinear correlation exists. In this case a regression equation of the form $y = a + bx$ defines the dependence of the phenomenon y on the phenomenon x . But for parabolic dependence the correlation coefficients merely define the prevailing general sense of the dependence, the sign + being used for direct dependence and the sign - (minus) for inverse dependence.

In our case, the measure of relationship is the correlation coefficient between the observed values y and the values y_1 computed from a parabolic regression equation. This correlation coefficient, which we designate by the symbol R_{y,y_1} , may be found by the ordinary method, but is far more simply defined by the formula:

$$R_{y,y_1} = \sqrt{\frac{\sum x^2 - \sum \beta^2}{\sum x^2}}$$

where $\sum x^2$ is the sum of the squares of the deviations of the values of y from their mean y_0 , while $\sum \beta^2$ is the sum of the squares of the deviations of the observed values y from the corresponding values y_1 , computed by the regression equation (1).

(1) The coefficient $R_{y,y}$ always has the plus sign, for the sense of the dependence is already taken into account by the regression equation.

In investigating the dependence of yield on rainfall, it must be remembered that, in general, its influence cannot be positive without limit or negative without limit, and that, according to climate, soil, subsoil, local relief, there is a point of optimum rainfall, on both sides of which the yield declines; the regression equation thus taking on the complex character of a parabolic dependence (1).

(1) I cannot go into great detail, in the present paper, on

the methods of determining the degree of a regression equation. I therefore limit myself to the general propositions which I hope to develop in the body of the work. 1. The fourth-degree equation

$$v = a + bx + cx^2 + dx^3 + ex^4$$

expressing the dependence of yield on rainfall, is sufficiently complete and general, for it is able: (a) to express the variation of productivity in the intervals of rainfall lying near its optimum norm; (b) to express the course of productivity as the conditions of vegetation deteriorate in consequence of decreasing rainfall; (c) to express the variation in productivity as the conditions of vegetation deteriorate in consequence of increasing rainfall. 2. We have the greatest chance of all for a fourth-degree equation in localities where the optimum rainfall coincides with its mean, normal value. As the mean and optimum values of the rainfall diverge, the equation has a tendency to reduce down to one of the third or second degree, and in the case where the optimum value goes outside the limits of the field of observation (which is possible in localities that have abundant rainfall, or those that suffer from insufficient moisture) it may reduce down to a linear equation. It is of course obvious that it is not only the amount of rainfall, but other meteorological elements as well (for example, temperature and humidity) and likewise soil conditions, influence the shifting of the optimum points. 3. Simplification of the regression line is also possible where the dependence is weak, since its detail movements, reproduced by the higher powers of x , may be lost among the random layers of superimposed variations. In this case there is the danger that a complex regression equation may transmit the random elements of the statistical material.

Investigation of the influence of rainfall must therefore be supplemented by location of the regression line and determination of the degree of its complexity. Analysis shows the dependence of yield on rainfall during the sowing period and of the final plowing for rye sowing to have a parabolic character, and to be expressed by [parabolic] regression equations. But for the period from germination to the onset of tillering, we must content ourselves with a linear equation, because of the danger that one of higher degree would transmit the random elements of the material (2).

(2) The analysis of the degree of the curves that express the character of the dependence between yield and rainfall had to be postponed for another paper, for the exposition of the methods employed by me would not find room in the framework of the present paper. I may say only that I passed to a more complex curve only when the results of the analysis dictated the need of such action. I acted in this way because of the danger that the curves would transmit the random variations of the statistical material. Thus, for instance, mathematical analysis showed that the dependence of yield on rainfall of the sowing season (and also on that of the period of the final plowing for rye seeding) did not fit a linear equation; whence the need for transition to a quadratic equation. But the same analysis revealed that further increase in the complexity of the curves might lead to their reproducing the random variations of the statistical material, for which reason I remained with my quadratic equations, though it might have been possible that the dependences, according to the corresponding to the periods under study, might better be reproduced by a cubic equation. In any case this scepticism cannot have important consequences: if analysis

reveals the danger that the randomness of the material will be reproduced by the regression equation, this is a sign that the detail variations marked by the third powers of p and q are small and are "engulfed" by random fluctuations. I have striven to protect the present tentative work from the reproach that the strong dependences obtained by it are merely a matter of chance -- and this is why I have preferred to be conservative in defining the degree of the regression equations.

Period of sowing: $y_p = 3.450 - 0.1022(p - 52) - 0.008775(p - 52)^2$ (II)

Period of final plowing: $p_m = 2.487 - 0.07672(p - 52) - 0.004401(p - 52)^2$ (III)

Period from germination

to inception of tillering: $y_p = -0.09357(p - 44)$ (IV)

We compare the indices of productivity y^* , calculated by the regression equation (II) with the observed productivity y^* (see Table 2).

* The former is written y in the text, the latter y (starred) is omitted. This is obviously a typographical error. See text, p. 263, second paragraph. Translator's note.

TABLE 2

Year	P_c	y	yP_c	Year	P_m	y	yP_m
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1905	90	- 16.4	- 13.4	1895	92	- 15.3	- 12.4
1894	89	- 11.9	- 12.6	1903	88	- 8.7	- 10.4
1895	87	- 15.3	- 11.1	1905	83	- 16.4	- 7.2
1897	83	+ 1.3	- 8.3	1893	80	- 5.6	- 5.8
1907	79	- 4.3	- 5.0	1887	74	+ 8.0	- 3.2
1913	76	- 2.1	- 4.2	1908	72	+ 8.7	- 2.4
1904	72	- 5.8	- 2.2	1911	68	+ 2.6	- 1.0
1884	71	- 7.2	- 1.7	1884	67	- 7.2	- 0.7
1912	67	+ 6.6	- 0.1	1890	63	+ 10.2	+ 0.4
1898	66	+ 4.0	+ 0.3	1907	58	- 4.3	+ 1.5
1906	64	- 4.8	+ 0.9	1909	58	- 2.1	+ 1.5
1902	63	- 6.4	+ 1.2	1907	55	- 5.8	+ 1.9
1900	61	+ 8.1	+ 1.8	1894	54	+ 1.3	+ 2.1
1909	55	- 2.1	+ 3.1	1910	54	+ 1.7	+ 2.1
1886	45	+ 4.1	+ 3.7	1906	53	- 4.8	+ 2.2
1911	45	+ 2.6	+ 3.7	1892	53	- 4.1	+ 2.2
1908	44	+ 8.7	+ 3.7	1891	52	+ 8.7	+ 2.3
1893	44	- 5.6	+ 3.7	1889	45	+ 2.9	+ 2.7
1916	43	- 3.3	+ 3.6	1886	41	+ 4.1	+ 2.6
1901	43	+ 7.7	+ 3.6	1901	41	+ 7.7	+ 2.6
1896	40	+ 8.1	+ 3.4	1898	41	+ 4.0	+ 2.6
1883	40	+ 0.1	+ 3.4	1888	39	+ 8.4	+ 2.5

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1903	39	- 8.7	+ 3.3	1883	38	+ 0.1	+ 2.5
1890	39	+ 10.2	+ 3.3	1894	37	- 11.9	+ 2.4
1891	39	+ 8.7	+ 3.3	1914	36	+ 5.4	+ 2.3
1915	39	+ 10.9	+ 3.3	1885	32	- 1.5	+ 1.7
1888	38	+ 8.4	+ 3.1	1896	32	+ 8.1	+ 1.7
1892	37	- 4.1	+ 3.0	1912	30	+ 6.6	+ 1.4
1914	35	+ 5.4	+ 2.6	1916	30	- 3.3	+ 1.4
1910	33	+ 1.7	+ 2.2	1915	29	+ 10.9	+ 1.2
1887	31	+ 8.0	+ 1.6	1899	26	- 9.0	+ 0.5
1889	26	+ 2.9	0.0	1902	23	- 6.4	- 0.3
1885	23	- 1.5	- 1.1	1900	22	+ 8.0	- 6.5
1899	23	- 9.0	- 1.1	1913	21	- 1.1	- 0.8

The indices of productivity for the sowing period ($y\rho_{\sim}$), the period of final plowing ($y\rho_{\sim}$), and the germination period ($y\rho_{\vee}$), for rainfall arranged by 5-millimeter-class-intervals, are as follows:

[See table on the following page]

The point of optimum rainfall, as may be seen from the annexed table, falls in the 45 millimeter interval and is appreciably shifted over to the dry side from the mean value or normal rainfall (52 millimeter) for the period under study. But while one branch of the parabola, tracing the course of productivity in dependence on the deterioration in sowing conditions that result from insufficient moisture, declines slightly and then quickly breaks off at the index -1.1, the other branch of the parabola, characterizing the

Rainfall

1	2	3	4	1	2	3	4
105	--	--	- 5.7	60	+ 2.1	+ 1.1	- 1.5
100	--	--	- 5.2	55	+ 3.1	+ 1.9	- 1.0
95	--	--	- 4.8	50	+ 3.6	+ 2.5	- 0.6
90	- 13.4	- 11.1	- 4.3	45	+ 3.7	+ 2.7	- 0.1
85	- 9.7	- 8.3	- 3.9	40	+ 3.4	+ 2.6	+ 0.4
80	- 6.4	- 5.8	- 3.4	35	+ 2.6	+ 2.1	+ 0.8
75	- 3.6	- 3.6	- 3.0	30	+ 1.3	+ 1.4	+ 1.3
70	- 1.3	- 1.7	- 2.5	25	- 0.4	+ 0.3	+ 1.8
65	+ 0.6	- 0.1	- 2.0	20	- 2.5	- 1.2	+ 2.3

variation in productivity in dependence on increasing rainfall, reacts weakly to the changing rainfall in the intervals near the optimum, and then gives progressively declining indices of productivity as the rainfall further increases. In general, the regression equation reproduces in very sharp relief the character of the dependence of yield in Petrograd guberniya on the rainfall in the sowing period; cases of harmful influence of insufficient moisture under the climatic conditions there are rarely observed and the dimensions of the damage are insignificant (two cases of rainfall under 25 millimeter, indices of productivity = 1.1), while the negative effect of abundant rainfall is far more frequently met and the damage

reaches considerable proportions (9 cases of rainfall over 67 millimeter, indices of productivity from -12.6 to -13.4 pads). The regression equation defining the course of productivity in dependence on the rainfall in the period of the final plowing before the seeding of rye, gives the same picture of the reaction of yield to rainfall, but only not so sharply expressed (the same optimum point at 45 millimeter, weak reaction of yield to variations in rainfall close to the optimum, insignificant effect of insufficient moisture, frequent cases of abundant rainfall and strong negative action thereof).

Let us now determine the measure of the rainfall influence in the sowing period, the final pre-sowing plowing period, and the period from germination to onset of tillering. As has been mentioned above, for this purpose we must find the correlation coefficients between the observed values y and the indices of yield, y^p_c , y^p_n , and y^p_v , computed from the regression equation,

$$\begin{aligned} R_{y, y^p_c} &= 0.65 \\ R_{y, y^p_n} &= 0.49 \\ R_{y, y^p_v} &= 0.23 (1) \end{aligned}$$

(1) For the case of rectilinear correlation, $R_{y, y^p_v} = R_{y^p_v}$

Thus the rainfall during the sowing period is reflected in the yield most strongly of all, while that in the pre-sowing period (i.e. of the last plowing before seeding the rye) is weaker, and that from germination to onset of tillering is most insignificant in effect.

Up to now we have been considering the isolated influence of the rainfall of the sowing and pre-sowing periods. But the action of the same amount of moisture during sowing time has a varying effect, according to the amount that has fallen during the pre-sowing period. Thus, for instance, the influence of rainy weather may be intensified, weakened or even completely paralyzed, depending on whether the rain in the pre-sowing period has been excessive or scanty. The same may also be said about the influence of drought at sowing time. This raises the question whether the periods of sowing and final plowing for sowing should be combined into a single more extensive period. In view of the fact that the influence of rainfall during sowing time is not equivalent to that of rainfall during the period of final plowing, a simple summation of the rainfall would be too simplified a method. These rains must be taken, to use a term of mathematical statistics, with different "weights". Moreover, no matter how feeble the influence of the rainfall in the germination period, it still introduces a certain modification into the condition of the winter sowings, and it is therefore appropriate to take it into consideration.

As a result we determine the rainfall of a single extended period from the beginning of the final plowing for rye seeding down to the onset of tillering of the winter rye, taking the "weights" as proportional to the respective correlation coefficients (2).

(2) I do not maintain that this is the best method of determining the "weights" in curvilinear correlation. A more perfect method is theoretically possible, but it is very complicated. Where the regression is rectilinear, the method of determining the weights

is very simple: namely, according to the coefficients of the regression equation defining y as a function of p_c, p_m, p_v .

Periods	R	Weights
Period of Sowing	0.65	0.47
Period of final plowing	0.49	0.36
Period of germination	0.23	<u>0.17</u>
Total		1.00

Designating the rainfall of this period by the symbol

we have:

$$p_{cnv} = 0.47p_c + 0.36p_m + 0.17p_v$$

The coefficient of correlation between the yield (y) and p_{cnv}

is:

$$r_{y, p_{cnv}} = -0.57$$

Thus the negative effect of the rainfall of the combined period is stronger than that of the component periods, though this should have been expected.

Further analysis shows that the dependence of the yield in this case will not fit into either a rectilinear or a second-degree regression equation, so that it becomes necessary to pass to an equation of the third degree:

$$y_{p_{cnv}} = 3.036 - 0.3945(p_{cnv} - 50) - 0.019034(p_{cnv} - 50)^2 + 0.00043861(p_{cnv} - 50)^3 \quad (V)$$

is very simple: namely, according to the coefficients of the regression equation defining y as a function of p_c, p_m, p_v .

Periods	R	Weights
Period of Sowing	0.65	0.47
Period of final plowing	0.49	0.36
Period of germination	0.23	<u>0.17</u>
Total		1.00

Designating the rainfall of this period by the symbol

we have:

$$p_{cnv} = 0.47p_c + 0.36p_m + 0.17p_v$$

The coefficient of correlation between the yield (y) and p_{cnv}

is:

$$r_{y, p_{cnv}} = -0.57$$

Thus the negative effect of the rainfall of the combined period is stronger than that of the component periods, though this should have been expected.

Further analysis shows that the dependence of the yield in this case will not fit into either a rectilinear or a second-degree regression equation, so that it becomes necessary to pass to an equation of the third degree:

$$y_{p_{cnv}} = 3.036 - 0.3995(p_{cnv} - 50) - 0.019034(p_{cnv} - 50)^2 + 0.00043861(p_{cnv} - 50)^3 \quad (V)$$

In Table 3 the observed productivity is compared with its indices computed by regression equation (V).

TABLE 3

Harvest	spg	y	y spg	Harvest	spg	y	y spg	spg: rainfall
Year				Year				Indices of pro- ductivity for rainfall grouped at 5-mm. group intervals
1895	85	-15.3	-15.5	1886	46	+4.1	+4.3	spg y spg
1905	79	-16.4	-13.9	1900	45	+8.1	+4.5	85 (-) 15.5
1894	70	-11.9	-9.1	1890	45	+10.2	+4.5	80 (-) 14.2
1897	69	+1.3	-8.4	1902	43	-6.4	+4.7	75 (-) 12.0
1884	67	+7.2	-7.1	1892	43	-4.1	+4.7	70 (-) 9.1
1907	66	-4.3	-6.4	1901	43	+7.1	+4.7	65 (-) 5.8
1913	61	-1.1	-3.1	1916	41	-3.3	+4.8	60 (-) 2.4
1904	59	-5.8	-1.8	1891	40	+8.7	+4.7	55 + 0.6
1909	58	-2.1	-1.2	1888	40	+8.4	+4.7	50 + 3.0
1903	57	-8.7	-0.5	1910	40	+1.7	+4.7	45 + 4.5
1906	57	-4.8	-0.5	1896	39	+8.1	-4.5	40 + 4.7
1893	56	-5.6	+0.1	1883	37	+0.1	+4.1	35 + 3.3
1898	54	+4.0	+1.2	1915	36	+10.9	+3.7	30 + 0.9
1912	53	+6.6	+1.6	1914	34	+5.4	+2.7	25 - 5.6
1908	52	+8.7	+2.2	1889	33	+2.9	+2.2	
1911	51	+2.6	+2.6	1899	27	-9.0	+3.2	
1887	48	+8.0	+3.8	1885	26	-1.5	-4.4	

It will be seen from the table that the unfavorable effect of heavy rainfall begins to appear from the 57 mm. interval. Of the 34 years covered by the observations, 11 had rainfall of 57 mm. and over [in this period], of which only one had a yield somewhat above normal

(+1.3). The productivity indices between 30 and 50 mm have the plus sign, with the optimum point falling in the 42 mm interval. Of the 21 years observed with rainfall in the limits of the favorable norms, only 4 give a yield under average, including in this number one year on the boundary of the abundant rainfall groups. Finally, inadequate moisture had a negative effect on the yield, under the climatic and soil conditions of Petrograd guberniya, in only 2 years of the entire 21 year interval. Diagram 1 graphically demonstrates the dependence of yield on the rainfall in the period from final plowing for the rye seeding down to the onset of tillering.

DIAGRAM 1

YIELD OF RYE AND THE RAINFALL DURING THE PERIOD FROM THE FINAL PLOWING
OF THE FINAL PLOWING TO THE ONSET OF TILLERING

Observed yield

Yield calculated by the regression equation:

$$y_{\text{spg}} = 3.036 - 0.3995(p_{\text{env}} - 50) - 0.01902(p_{\text{env}} - 50)^2 + 0.00043661(p_{\text{env}} - 50)^3.$$

Yield in Puds per
Desyatina, Expressed
as Deviations from
Norm

Rainfall in mm

The correlation coefficient between the observed productivity and its indices y mark the strong influence of the rainfall in the period under study on the rye yield in Petrograd guberniya:

$$R_{y,y}^{pny} = 0.74 \quad (1)$$

$$R_{y,y}^{pen} = 0.75$$

(1) If only the period of the final plowing and the time from sowing to germination are included in the combined period, we get almost an identical regression equation, and almost an identical correlation coefficient: R :

$$y_{pen} = 2.967 - 0.3408(p_{en} - 51) - 0.01667(p_{en} - 51)^2 + 0.00026026(p_{en} - 51)^3$$

Analysis of the conditions of the pre-sowing and sowing periods give results which allow us to maintain that the rainfall during the period under study is one of the principal factors for the yield of rye in Petrograd guberniya, with the negative effect of abundant rainfall standing out with exceptional clarity. The inverse dependence between the yield of rye in Petrograd guberniya on the rainfall was already noted, in its broad outlines, by A. F. Fortunatov in his classic work "The Rye Crops of European Russia". The guberniya studied (like that of Moscow) belongs to a belt of Russia which is distinguished for its rainy climate. According to observations of meteorological stations, the balance of influx of water influx [by rainfall] and outgo by way of evaporation has the following norms:

Stations	Rainfall	Evaporation	Balance
Pavlovsk (Petrograd)	583	305	+ 278
Petrograd	511	302	+ 209
Petrovsko-Razumovskoye (Moscow)	536	377	+ 159
Poltava Experimental Field	462	753	- 291
Lugansk	406	831	- 365
Astrakhan	156	721	- 565

(Cited in work of Vlasov "Climatology of Moscow guberniya").

The negative influence of abundant rainfall during the sowing period and the period of final plowing that immediately precedes is apparently explained mainly by the disturbance to the aeration of the soil. The water-soaked soil obstructs the access of air to the germinating grain, while the latter needs oxygen, since the processes of germination are accompanied by the energetic absorption of oxygen and the excretion of carbon dioxide. The rapid and marked increase in the negative effect of rainfall during the period of the final plowing and sowing, and its similarly rapid and strong decline from the moment of germination, leads to this explanation. The observations of the correspondents of the current agricultural statistics of the Petrograd guberniya remark on the weak germinating power of the seed, its rotting, and the stunted, weak seedlings, in the years when the weather is rainy during the time of sowing and final plowing. A summary of the reports on the relative germinating capacity of the seed according to the degree of dampness of the weather gives the following figures.

5/6

Rainfall of the third July Decade and August
in Millimeter

Figure 5 years with heaviest rainfall: 124-182 millimeter
4 years with rainfall of 103-123 millimeter
4 years with rainfall of 71-97 millimeter
4 years with scantiest rainfall: 61-70 millimeter

36.5
6.5
5.8
5.5

It should also be noted that water-soaked soil is difficult to cultivate and the quality of the cultivation suffers badly as a result.

Rainfall of the third July decade and the first two August decades	Percent of reports that land was badly prepared for seeding
---	--

5 years with heaviest rainfall: 91-119 millimeter	17.4
5 years with rainfall of 74-89 millimeter	8.2
5 years with rainfall of 52-73 millimeter	5.0
5 years with scantiest rainfall: 36-42 millimeter	5.0

The correspondents remark on the deterioration in the quality of the work during sowing time as well, as a result of rainy weather, but there are no numerical tabulations in the "Collection" for Petrograd guberniya.

Reference to these conditions does not exclude the negative actions of other factors. Thus, for example, the abundant rainfall during the period of the final plowing and the sowing are reflected unfavorably in the soil processes, and possible even more strongly

than at other times since the plowed and harrowed soil has had its vegetative cover removed and the nutrient salts washed out of it (principally nitrites) are not taken up by the roots of the plant. But this is important for the future plants: in the period of the "underground" existence, the plant is nourished by the elements stored up in the seed, and does not require the substances found in the soil.

The soil processes apparently have their primary if not their exclusive influence during the fallow period of the field to be seeded. As we have seen, the first two July decades, even though to a very slight extent, still have a negative influence on the yield of rye. Going back still further to June and May of the year for the harvest, we determine the following correlation coefficients between yield and rainfall:

First two July decades of the year before the harvest	$r = -0.15$
All of June of the year before the harvest	$r = -0.33$
All of May of the year before the harvest	$r = -0.40$

If we designate the aggregate rainfall of May, June and the first two decades of the year before the harvest by the symbol m^p , we find the correlation coefficient $x y$ (yield) and m^p

$$r_{m^p} = -0.46$$

The correlation coefficient is 5 times its probable error, and therefore its origin may be ascribed to chance only by the wildest stretch of the imagination.

The corresponding regression equation is expressed by the formula:

$$y_{m^p} = -0.0919(m^p - 145)$$

Analysis of the character of the dependence of yield on the rainfall of the fallow period shows that, under the climatic and soil conditions of Petrograd guberniya, it is expressed by a rectilinear regression, and therefore there is no need for a parabolic curve. The circumstance that the correlation coefficients do not increase as the sowing time approaches compels us to assume that the center of gravity of the negative effect resides in the soil processes rather than in the strengthening or weakening of the action of the rainfall during sowing time (1).

(1) The correlation coefficients computed separately for May, June, and the first two decades of July, not only fail to rise, but even, regularly and strongly, fall as the seeding time approaches. In the present state of this work, there is no possibility of connecting this fall with the soil processes. Therefore I cannot decide whether this fall is an actual phenomenon or merely a matter of chance. Only after a detailed work-up of the material by short ten-day intervals, in connection with the work on a fallow field (spreading and plowing in manure, plowing and harrowing of the field) can we reach any definite conclusion. In view of the doubtful existence of this fall in the correlation coefficient, I did not decide to place the aggregate rainfall of the fallowing period in dependence on the degree of influence of the May, June and July rainfall* but confined myself to its simple summation. It is appropriate at this point to note in somewhat greater detail the harmful influence of excessive rainfall on the soil processes. 1. We have already pointed out that abundant moistening of the soil leads to depletion of its fertility, since the nutrient salts are washed out of it. This process of defertilization is strengthened by the formation of humic acids, which

break up the mineral parts of the soil and carry the potassium and phosphorus salts into the subsoil and the ground waters. 2. In low places with a tendency to waterlogging, excessive rainfall, in connection with poor aeration of the soil, leads to the formation of ferrous oxide and sulfide, which are injurious to plants. 3. Excessive soil moisture disturbs bacterial activity in it, which activity is immensely important in the nutrient regime of the plant. Water-soaked soil becomes impermeable to air. As a result the beneficial processes of aerobic decomposition of organic residues are paralyzed and conditions created for the unfavorable processes of anaerobic decomposition of organic substances instead. (Cf. "General Agriculture" of Professor V. R. Vilyams, pp. 70-73, 369-371, on the significance of aerobic and anaerobic processes).]

* This is apparently an obscure reference to the author's method of weighting the rain fall for the separate periods, proportionally to the correlation coefficients. Cf. p. 269 of text. [Translator's Note].

We now take up the study of the dependence of yield on the rainfall of the third September decade, i. e. before the onset of tillering in winter rye.

The correlation coefficient between yield and rainfall in this decade takes a strong jump and changes from negative to positive.

$$r_{y/x} = +0.34$$

The correlation coefficient for October shows a very weak positive reaction of yield to rainfall, completely within the range of random fluctuations.

$$r_{y,p} = +0.10$$

We are consequently justified in drawing the following conclusion:

The period of the inception of tillering is characterized by a direct dependence of yield of rainfall; subsequently this dependence is either sharply reduced or vanishes entirely (1).

(1) My material for the October decades only covers a 22-year period. It definitely indicates a sharp fall in the positive reaction of yield to rainfall, from the first October decade.

The period when tillering begins is the only period in which a positive reaction of yield to rainfall is noted for Petrograd guberniya. Therefore the doubt as to whether the index of this dependence (a correlation coefficient of +0.34) is not the result of a random combination of factors is entirely natural. Let us test this inference by the methods of mathematical statistics.

(1) The correlation coefficient of +0.34 for 34 observations is 3.3 times its standard error of 0.10225. There are therefore 12 chances out of 1,000 that the coefficient is due to chance.

(2) We divide the 34 years into two groups, placing in one group the 17 years with highest rainfall in the onset-of-tillering period, and in the other group the 17 years with lowest rainfall in that period. 13 of the 17 years in the first group have a yield

above the norm, and 4 years are below it. 5 of the years in the second group have a yield over average, and 12 have a yield under average.

Consequently the frequency of good crop years in the first group is 0.76

Consequently the frequency of good crop years in the second group is 0.29

Divergence 0.47

The divergence between these frequencies is 4.5 times the probable error [VEROYATNOST ODKLONIENIYE] of the difference, which is equal to 0.103. There are thus 12 chances in 10,000 that the difference is due to chance.

(3) In the first group of years the mean yield exceeds the	
norm by	+ 3.4 puds
In the second group of years the mean yield exceeds	
the norm by	- 3.4 puds

The divergence of the observed means is 6.8 puds (+ 3.4 puds - (- 3.4)); the standard error of difference is 1.5 puds. Consequently the observed difference is 4.5 times its standard error. There are 12 chances in 10,000 that the difference is due to chance.

The general conclusion that follows from this analysis is as follows: the positive reaction of the rye yield to the rainfall during the period of the onset of tillering, which is observed in Petrograd guberniya, can only be ascribed to chance by an exceedingly great stretch of the imagination.

Let us try to assign the most probable explanation for this

dependence. The period of the inception of tillering is characterized by: (1) intensified formation of the plant-body, its overground and underground parts; (2) shallow location of the roots, since the root system is only beginning to develop (the tillering period is at the same time a period of taking root). Energetic growth should be accompanied by intensified nutrition of the plant, which absorbs mineral salts from the soil in the form of dilute aqueous solutions. Since the root system is only slightly developed at the onset of tillering and lies at shallow levels, abundant moistening of the surface layers of the soil is a condition of good nutrition. As the root system develops and goes into the deeper -- and therefore moister -- soil layers, plant nutrition becomes more independent of the moisture conditions in the upper layers. Moreover, as the roots go deeper, the access of the air they require for respiration becomes more and more difficult, and since abundant rainfall (especially under the soil conditions of Petrograd guberniya) interferes with the process of soil aeration, a negative effect of rainfall (1) begins to be observed, parallel to its positive effect.

(1) As we shall see, for the spring and summer periods the negative effect of rainfall outweighs its positive action.

The regression equation defining the dependence of yield on the rainfall at the onset of tillering (3rd September decade) is as follows:

$$y_{p \frac{3}{ix}} = +0.02 + 0.2307 \left(p_{\frac{3}{ix}} - 17 \right) \quad \text{VI}$$

Analysis shows that the transition to the second-degree equation:

$$2y_{p \frac{3}{ix}} = 1.114 + 0.2622 \left(p_{\frac{3}{ix}} - 17 \right) - 0.009217 \left(p_{\frac{3}{ix}} - 17 \right)^2 \quad \text{VII}$$

implies the danger of reproducing the random fluctuations of the statistical material (along with the real detail), and it is thus preferable to remain with the linear equation. The table below gives two variants of the indices of productivity in dependence on the rainfall of the third September decade.

TABLE 1

Rainfall $\frac{\rho_3}{ix}$	$1y \frac{\rho_3}{ix}$	$2y \frac{\rho_3}{ix}$
1	2	3
40	+ 5.3	+ 2.3
35	+ 4.1	+ 2.8
30	+ 3.0	+ 2.9
25	+ 1.8	+ 2.6
20	+ 0.7	+ 1.8
15	- 0.4	+ 0.6
10	- 1.6	- 1.2
5	- 2.7	- 3.4
0	- 3.8	- 6.0

We have thus considered the influence of rainfall on the yield of rye, beginning with May and ending with October (of the year preceding the harvest), i. e. during the following period, the sowing period and the two-month period of the autumn life of winter rye. We pass now to the temperature conditions.

The correlation coefficients between yield and temperature in August, September, and October are:

August temperature	$-1\bar{y}, t_{VIII}$	$= + 0.06$
September temperature	$-1\bar{y}, t_{IX}$	$= + 0.28$
October temperature	$-1\bar{y}, t_{X}$	$= + 0.15$

The correlation for August is close to zero, but then this was to be expected, since for the first two decades of August the field is not occupied by sowings, and reaction of yield to temperature is thus possible only from the third August decade on. The beneficent influence of warmth on the growth of winter rye is characterized by the positive correlation coefficients for September and October. But these coefficients (especially for October) are small.

The positive reaction of yield to temperature is outlined far more definitely when the temperature conditions are more precisely arranged by their incidence on the separate vegetative phases, by short ten-day intervals (1).

(1) This could have been done for the 22-year harvest period from 1895 to 1916, for which statistical data is available on each year of the period. The average time of intensified sowing, as has been shown, fell on 25 August New Style, but on some years varied between 22 August and 3 September. The 10-day periods given below were computed for each year individually depending on the date of sowing.

DEPENDENCE OF YIELD ON TEMPERATURE AT TEN-DAY INTERVALS (OVER 22-YEAR PERIOD)

2nd decade before sowing	$r_{y,-2} = + 0.11$
1st decade before sowing	$r_{y,-1} = 0.00$
Decade of sowing	$r_{y,0} = + 0.29$
1st decade after sowing	$r_{y,1} = + 0.37$
2nd decade after sowing	$r_{y,2} = + 0.41$
3rd decade after sowing	$r_{y,3} = + 0.33$
4th decade after sowing	$r_{y,4} = + 0.34$
5th decade after sowing	$r_{y,5} = + 0.49$
6th decade after sowing	$r_{y,6} = + 0.43$
7th decade after sowing	$r_{y,7} = + 0.41$
8th decade after sowing	$r_{y,8} = + 0.28$

There is no reaction of yield to temperature in the first two decades before sowing. The correlation coefficient jumps sharply to + 0.29 in the sowing period. As is commonly known, warmth is a factor that favors the germination of grain. But since, on the one hand, rye also germinates at low temperatures (even at + 1.5), and, on the other hand, the weather is usually warm during the sowing period (its mean temperature for the 22-year period was + 14.2 fluctuating during that time from + 11.9 to + 16.8 in the various years), no strong reaction of yield to the temperature at sowing time is observed. The emerging seedlings require a still larger amount of warmth for their growth, and the temperature during the growth period of winter crops falls continuously (1), for which reason its positive effect is more strongly manifested: the correlation coefficients from the 1st to 7th week decades after sowing, inclusive, range from + 0.33 to + 0.49, and show no definite tendency to rise or fall. This is confirmed by combining the decades into 20-day periods,

and then into monthly periods.

(1) The successive mean temperatures of the first six decades are:
 $+12.9; +10.4; +9.0; +6.8; +4.9; +2.7.$

Sum of mean 10-day temperatures

Correlation coefficients

1st and 2nd decades after sowing

$$r_{y, 1+2} = +0.44$$

3rd and 4th decades after sowing

$$r_{y, 3+4} = +0.37$$

5th and 6th decades after sowing

$$r_{y, 5+6} = +0.50$$

1st, 2nd and 3rd decades after sowing

$$r_{y, 1+2+3} = +0.50$$

4th, 5th and 6th decades after sowing

$$r_{y, 4+5+6} = +0.49$$

If we denote the sum of the 10-day temperatures of the first two months after sowing, we find:

$$r_{y, t_2} = +0.55$$

Finally, on summing the mean temperatures of the decades of the entire autumn life of winter rye, i.e. from the moment of germination to the onset of freezing temperatures (2), we determine the correlation coefficient between y and t_0 (the sum of the 10-day temperatures) to be:

$$r_{y, t_0} = +0.60$$

(2) Up to the decade with mean temperature below zero.

The available material allows us to solve the meteorological question of the significance attaching to the accurate assignment of temperatures to the several vegetative stages of the autumnal life of winter rye. The 10-day, 20-day and monthly correlation coefficients are computed from the material for the 22 years for which we

have information on the sowing times for each year. The average date of sowing for the 22 years is 25 August, and consequently the average date of germination is 31 August. Thus on the average the temperatures in September and October are, on the average, the temperatures of the first and second months of growth of the winter rye. We find the coefficients of correlation between the yield and the September and October temperatures, and compare them with the correlation coefficients computed with a more exact regard for the vegetative periods.

Periods of Growth	Correlation Coefficients	
	With precise consideration for the vegetative periods	For September and October
First month of growth of winter rye	+ 0.49	+ 0.43
Second month of growth of winter rye	+ 0.50	+ 0.31
For the two-month period	+ 0.55	+ 0.40

In spite of the fact that the fluctuation of the sowing time is not particularly great from year to year, and lies within the limits of 12 days (22 August to 3 September), the significance of the more precise method stands out very sharply, and the degree to which the autumn temperature conditions influence the yield of rye is reflected far more definitely by it.

Unfortunately information on the sowing time is not available for each of the 34 years covered by this study. We must therefore

limit our comparison of yield with the temperature conditions of September and October, but it must not be forgotten that the reaction of yield to autumn temperature is reproduced only incompletely and to an attenuated degree by the regression equation:

$$14t_{IX,X} = 2,010(t_{IX,X} - 8.9) \quad \text{VIII}$$

where $t_{IX,X}$ = the sum of September and October temperatures, each of them weighted ($t_{IX,X} = 0.76$ September temperature + 0.24 October temperature) (1).

(1) The "weights" of the temperatures are fixed according to the coefficients of the September and October temperatures in the regression equation that defines y as a function of these temperatures.

The correlation coefficient between y and $t_{IX,X}$ is:

$$14r_{y,t_{IX,X}} = +0.30 \quad (2)$$

(2) In fact the coefficient between y and $t_{IX,X}$ would be in all probability equal to $+0.40$, $+0.45$, if the vegetative periods were taken into account for each year. And the coefficient of correlation with the aggregate temperatures from germination to the onset of freezing weather lies between $+0.45$ and $+0.50$.

The influence of the temperature conditions in the year preceding the harvest is not confined to the period we have just considered. The correlations established for the April, May, June and July temperatures reveal a positive influence strange at first glance, of the July temperature on the yield of the following year, and the almost complete neutrality of the temperatures for the other months:

Temperatures in the Year Preceding the Harvest	Correlation Coefficient with the Yield
April	+ 0.12
May	+ 0.13
June	+ 0.03
July	+ 0.40
August	+ 0.06

The correlation between July temperature and yield is + 0.40 and is four times its probable error.

$$\frac{0.40}{0.0995} = 4.05$$

There are thus 34 chances in 10,000 that it is of chance origin. The following explanation of the positive effect of July temperature on the yield of rye in the following year is very plausible. The rye harvest in Petrograd guberniya begins on the average on 2 August, New Style. The maturation of rye therefore usually occurs in July. High temperature during this period gives a fully ripened grain, while low temperatures acts unfavorably on its degree of ripeness. The stage of ripeness attained by grain is one of the elements of its quality for seeding purposes. Let us test this hypothesis by the same method, already tried, of short ten-day periods, applied precisely to the moment that harvest begins in each of the 22 years for which this information is available (1):

(1) The fluctuations of the time of the commencement of harvest in Petrograd Guberniya are very considerable: the earliest maturation

of rye was observed in 1896 (23 June New Style), the latest in 1902 (31 August New Style).

5th decade before the commencement of harvesting	$r_{y,-5} = -0.06$
4th decade before the commencement of harvesting	$r_{y,-4} = +0.30$
3rd decade before the commencement of harvesting	$r_{y,-3} = +0.48$
2nd decade before the commencement of harvesting	$r_{y,-2} = +0.25$
1st decade before the commencement of harvesting	$r_{y,-1} = +0.33$
1st decade after the commencement of harvesting	$r_{y,+1} = +0.13$
2nd decade after the commencement of harvesting	$r_{y,+2} = +0.07$

The correlation coefficient between the yield of the following year and the temperature jumps sharply in the fourth decade before harvesting, and holds within the limits of $+0.25$ and $+0.48$ throughout the 40-day period, to drop strongly at the moment of maturation of the rye. These limits of the positive effect of the temperature during the 40-day period before the harvest begins stand out still more sharply if the 10-day intervals are combined into 20-day intervals:

5th decade before harvesting	$r_{y,-5} = -0.06$
4th and 3rd decades before harvesting	$r_{y,-(4+3)} = +0.54$
2nd and 1st decades before harvesting	$r_{y,-(2+1)} = +0.35$
1st and 2nd decades after harvesting	$r_{y, 1+2} = +0.10$

The correlation coefficient between the yield of the following year and the sum of the 10-day [mean] temperatures for the 40-day period before the harvest is:

$$-4r_{y,t} = +0.55$$

The correlation coefficient between the yield and the July temperature of the prior year is $+0.46$ for the same 22 years.

When the question is worked up by 10 day intervals, applying them to the beginning of the harvest in each year, not only confirms the hypothesis set forth above but allows us to expand its content. It develops that the temperature conditions during the entire period of grain formation, beginning with blooming and ending with maturation, is beneficially reflected in the yield of the coming year (1). Obviously high temperature during the period under study improves the quality of the grain as seeding material, while a low temperature reduces that quality. Thus the connecting element between the temperature conditions of the 40 day period before harvesting, and the yield of the following year, is the grain used as seeding material.

(1) Rye usually begins to bloom in Petrograd guberniya on the 42nd day before maturation.

Apparently the quality of the seeding material is determined not so much by the degree of ripeness of the grain as by other conditions, as a result of the temperature during the 40 day period of grain formation. At least the continuous fall in the positive effect of temperature as the harvest time approaches leads to this assumption.

4th and 3rd decades before harvest	$r = +0.54$
3rd and 2nd decades before harvest	$r = +0.42$
2nd and 1st decades before harvest	$r = +0.35$

As is well known, a high temperature is one of the factors in the favorable course of the blooming process. This is not surprising,

because under favorable circumstances of germination, a viable "progeny" is obtained. It is possible, moreover, that the chemical composition of the grain, which depends on temperature conditions, is not without influence on its quality as seeding material. As is well known, high temperature increases the percentage of nitrogen in grain.

The regression equation that reproduces the dependence of yield on the temperature of the July in the preceding year (see equation IX below), of course, characterizes this dependence as being weaker than it really is, since, it is impossible in the first place to apply the periods precisely to the time when harvesting commences, and in the second place it was necessary to confine the equation to 3 decades instead of 4:

$$y_{-1 \pm VII} = 1.718(t_{-1 \pm VII} - 17.7) \quad (IX)$$

For the sake of completeness of this study, I still note the influence on the yield of the following year exerted by the nature of the grain, which is one of the essential elements determining the quality of the seeding material. On comparing the weight of a quarter of rye with the yield of the following year, according to the data of the central statistical committee for a 28-year period, I found the following coefficient of correlation:

$$r_{y,w} = +0.31 \quad (2)$$

(2) The data of the Central Statistical Committee give the nature of the grain in fairly rough indices: in the weight of a quarter in puds, correct to one decimal place. This coarseness of method, in all probability, is reflected in too low a value for the coefficient of correlation.

because under favorable circumstances of germination, a viable "progeny" is obtained. It is possible, moreover, that the chemical composition of the grain, which depends on temperature conditions, is not without influence on its quality as seeding material. As is well known, high temperature increases the percentage of nitrogen in grain.

The regression equation that reproduces the dependence of yield on the temperature of the July in the preceding year (see equation IX below), of course, characterizes this dependence as being weaker than it really is, since, it is impossible in the first place to apply the periods precisely to the time when harvesting commences, and in the second place it was necessary to confine the equation to 3 decades instead of 4:

$$y_{-1 \pm VII} = 1.718 (-1 \pm VII - 17.7) \quad (IX)$$

For the sake of completeness of this study, I still note the influence on the yield of the following year exerted by the nature of the grain, which is one of the essential elements determining the quality of the seeding material. On comparing the weight of a quarter of rye with the yield of the following year, according to the data of the central statistical committee for a 28-year period, I found the following coefficient of correlation:

$$r_{y,w} = + 0.31 \quad (2)$$

(2) The data of the Central Statistical Committee give the nature of the grain in fairly rough indices: in the weight of a quarter in puds, correct to one decimal place. This coarseness of method, in all probability, is reflected in too low a value for the coefficient of correlation.

III.

I turn now to the consideration of the influence of the meteorological conditions of the spring and summer life of winter rye on the yield of rye in Petrograd guberniya. The coefficients of correlation between the yield and the May, June, and July rainfall are as follows

May rainfall	$r_{y,pV} = -0.28$
June rainfall	$r_{y,pVI} = -0.50$
July rainfall	$r_{y,pVII} = -0.36$

The combined influence of the May, June, and July rainfall is defined by the following regression equation:

$$Y_{pV,VI,VII} = +0.109 - 0.0413(pV - 44) - 0.1615(pVI - 57) - 0.0894(pVII - 70) \quad (\bar{X})$$

(10)

As we see, the rainfall during the 3 month period of the spring life of winter rye has a negative effect on the yield of rye, with the June rainfall standing out in the strength of that effect.

The coefficients for the May, June, and July rainfall constitute a measure of the "weights" of the rainfall to be used when these months are combined to a single period:

	Coefficient of the equation "Weights"	
May rainfall	0.0413	0.14
June rainfall	0.1615	0.55
July rainfall	0.0894	<u>0.31</u>
Total		1.00

On summing the rainfall of May, June, and July, taken with the corresponding "weight" for each month, and denoting the amount thus determined by the symbol $P_{V,VI,VII}$, we find the correlation coefficient between p and the yield to be:

$$r_{(V,VI,VII)} = -0.61; (p_{V,VI,VII} = 0.14p_V + 0.55p_{VI} + 0.31p_{VII}),$$

(1)

While the corresponding regression equation has the following form

$$Y_{p_{V,VI,VII}} = -0.147 - 0.293(p_{V,VI,VII} - 60) \quad (XI)$$

(2)

The indices of productivity, $Y_{p_{V,VI,VII}}$, are compared in table I with the observed yield.

Mathematical analysis reveals no signs of curvilinear regression, and for this reason equation (XI) satisfactorily expresses the dependence of yield on the rainfall during the period of the spring and summer life of rye. Thus under the climatic and soil condition of Petrograd guberniya, during the period under study the rainfall exerts a negative influence on the yield of rye: the less rain falls, the better the yield, and conversely, rainy weather results in a poor rye harvest. Within the limits of the field of observation (that is during the 34-year period) no optimum rainfall points are observed, on both

sides of which the productivity declines. This inverse character of the dependence is so sharply and clearly expressed, that complex method of mathematical statistics are not needed to note its essential feature: the 6 years with the highest rainfall, 71 mm. and over, all have yields under average, while on the other hand the 8 driest years with 47 mm. and under, all gave yields of rye above the average/1/

/1/ It is possible that a more scrupulous investigation in the future will reveal a certain curvilinearity of regression for the rainfall of this period as well. But even now we are in a position to say that this suppositious regression line would in any case be little different from a straight line, for if there were any appreciable difference from a straight line it would have been noted in any case in the present work. The negative effect of abundant rainfall finds its explanation, in the first place, in soil processes, especially in the washing out from the soil of nitrogen salts; and in the second place in the inefficient aeration of the soil: the excessive moistening of the soil impedes the access of air to the plant roots, which during this period of plant life are already deep, and thereby it disturbs the process of root respiration; thirdly, by the specific action of rainy weather during the various periods; thus, for instance, rainy weather affects the process of blooming harmfully. While extreme raininess during the period of ripening and maturation occasions lodging of the cereal crops, as a result of which the plants undergo unfavorable conditions of ripening and maturation and this in addition creates the danger of root-rot.

The absence of any harmful influence caused by insufficient moisture in Petrograd guberniya appears far more strange: the yield of rye always increases as the rainfall declines (see diagram 2). Under the climatic conditions of Petrograd guberniya (by which we understand the totality of conditions: the amount of rainfall, the atmospheric humidity, a relatively moderate temperature) there are apparently only seldom droughts which are harmful for rye, and when they do come, they are only of short duration, and the plant, with the development by this period of a root system that penetrates deeply into the lower layers of the soil, which are better supplied with moisture, easily survives it.

Moreover a certain relation of inverse character is observed between the amount of rainfall and the relative yield of rye grain to 100 puds of straw (in Petrograd guberniya an average of 51 puds of grain are harvested for each 100 pounds of straw, and this relative grain yield fluctuates within the range of 37 to 63 puds);

$$\begin{aligned} r_{q,pV} &= -0.60 & q \text{ is the relative grain yield per 100 puds} \\ & & \text{of straw} \\ r_{q,pVI} &= -0.64 & pV \text{ is the May rainfall} \\ r_{q,pVII} &= -0.36 & pVI \text{ is the June rainfall} \\ & & pVII \text{ is the July rainfall} \end{aligned}$$

where q is the relative grain yield per 100 puds of straw, and are the rainfall in May, June and July respectively.

The regression equation is

$$q_1 = 0.35 - 0.1312(pV - 44) - 0.1330(pVI - 57) - 0.0429(pVII - 70).$$

(1)

Table 1

Year of Harvest	$\rho_{V,VI,VII}$	y	$\rho_{V,VI,VII}$	Year of Harvest	$\rho_{V,VI,VII}$	y	$\rho_{V,VI,VII}$	
1	2	3	4	1	2	3	4	
1894	107	-11.9	-13.4	1890	56	+10.2	+1.0	$\rho_{V,VI,VII}$: rainfall of May June, and July, taken with weights y: observed yield
1892	87	-4.1	-8.1	1898	56	+4.0	+1.0	
1902	83	-6.4	-6.9	1885	55	-1.5	+1.3	
1906	80	-4.8	-6.0	1915	53	+10.9	+1.9	y $\rho_{V,VI,VII}$: /computed/ productivity indices
1895	74	-15.3	-4.3	1907	53	-4.3	+1.9	
1916	71	-3.3	-3.4	1913	53	-1.1	+1.9	
1897	69	+1.3	-2.8	1901	52	+7.7	+2.2	
1883	68	+0.1	-2.5	1886	52	+4.1	+2.2	
1887	68	+8.0	-2.5	1909	50	-2.1	+2.8	
1903	67	-8.7	-2.2	1896	47	+8.1	+3.7	
1904	67	-5.8	-2.2	1912	45	+6.6	+4.3	
1905	67	-16.4	-2.2	1910	44	+1.7	+4.5	
1908	67	+8.7	-2.2	1900	43	+8.1	+4.8	
1899	63	-9.0	-1.0	1914	41	+5.4	+5.4	
1884	60	-7.2	-0.1	1888	38	+8.4	+6.3	
1911	60	+2.6	-0.1	1889	36	+2.9	+6.9	
1893	57	-5.6	+0.7	1891	34	+8.7	+7.5	

The complete coefficient of correlation between q and $P_{V,VI,VII}$, where $P_{V,VI,VII}$ is the aggregate rainfall for all three months, taken with weights, is

$$r_{q,P_{V,VI,VII}} = -0.79; \text{ and } P_{V,VI,VII} = 0.43 P_V + 0.43 P_{VI} + 0.14 P_{VII}$$

the weights being taken proportionate to the corresponding coefficients in the regression equation.

Thus in rainy years less grain is harvested than in dry years and the negative effect of the May and June rainfall /in all years/ is far stronger than that of the July rainfall, in spite of the fact that it is precisely during the month of July that the process of grain formation -- its ripening and maturation -- takes place.

A partial explanation for June may be found in the harmful of rainy weather during the blooming period, i.e. during the incipient period of grain formation. During May, however, no processes connected with grain formation are observed. The stronger negative effect of May rain on grain formation in comparison to July rainfall therefore requires other hypotheses. Attention must apparently be turned to the soil processes. Up to the moment of blooming, all the necessary elements for grain formation, taken from the soil, including nitrogen salts, are contained in the body of the plant. The straw differs in chemical composition from the grain: the latter contains relatively more nitrogen salts but relatively less potassium compounds than the straw. Since the nitrogen salts are more soluble in rainy years, they are more easily washed out of the soil than the potassium salts. Consequently, in rainy years the elements that accumulate in the body of the plant are more favorable to the formation of straw than to that

of grain. On the other hand, in dry years more nitrogen accumulates in the body of the plant, and this favors the formation of grain. But this hypothesis, of course, needs further verification. It may be ultimately modified, supplemented or even rejected entirely.

Thus there was not a single year, during the 34-year period, in which a poor harvest of rye could be attributed to insufficient rainfall during the time of the spring and summer life of the plant. The method of graphic representation confirms these conclusions. Diagram 2 shows firstly the observed productivity and secondly the regression line determined by Equation XI and giving the "theoretical" indices of productivity. The broken line of observed productivity has a tendency to uninterrupted growth with declining rainfall, and the line of regression (by which one may be convinced by ocular demonstration) very satisfactorily reproduces this tendency.

The study of the dependence of yield on temperature conditions gives results which are, on first glance, unexpected. The correlation coefficients for May and June are marked by the absence of any relationship between yield and the amount of heat during these months, but the coefficient for July shows a weak positive reaction of yield to July temperature:

May temperature	$r = - 0.02$
June temperature	$r = - 0.07$
July temperature	$r = + 0.26$

The "neutrality" of the May and June temperatures is in contradiction to the customary ideas of the local inhabitants on the significance of warmth during these months for the yield of rye and requires

shorter, in cold years longer /2/

/2/ In Moscow guberniya rye blooms as early as 16 May, Old Calendar, in warm years; in cold years as late as 17 June. From the scanty amount of information available for Petrograd guberniya, a year of very early blooming may be found (1898, on 20 May) and a year of late blooming too (1902, at the end of June, Old Calendar). To go by the observations of the Moscow Statistical Bureau alone, the blooming period is almost unaffected by the weather and lasts about 14 days.

Thus the aggregate amount of heat /during the blooming period/ is evened out to a considerable extent. Moreover, the warmer month of June follows May, and the still warmer month of July follows June. Consequently a low temperature in May always finds its "antidote" in the higher June temperature, as does a low June temperature in the higher July temperature. But a low temperature in July (during which, on the average, the maturation period of rye falls) does not always find its compensation in a higher August temperature, since August is marked by a fall in temperature (from 17.6 degrees in July to 15.2 degrees in August). Therefore the danger of unfavorable maturation conditions is a threat in years of low July temperature. We recall that a steady fall in the temperature is similarly observed during the autumn life of winter cereals, which period is likewise characterized by the positive effect of warmth on the yield. Thus during the spring and summer life of rye the rainfall of May, June and July exert their influence, and, to a lesser extent, the temperature of July. Let us determine the combined influence of all these factors

explanation /1/

/1/ It should be noted that A.F. Fortunatov, in his "Rye Crop of European Russia", already came to the conclusion that this idea was not confirmed by statistical analysis in the region to which Petrograd guberniya belongs.

Diagram 2

YIELD OF RYE AND THE SPRING AND SUMMER RAINS IN PETROGRAD GUBERNIYA

Observed yield

Yield computed from the regression equations

$$\bar{p}_{V,VI,VII} = -0.147 - 0.2931 (\bar{p}_{V,VI,VII} - 60)$$

YIELD IN PUDS PER DESYATIN, EXPRESSED

AS DEVIATIONS FROM THE MEANS

+ 15
+ 10
+ 5
0
- 5
- 10
- 15
- 20

105 100 95 90 85 80 75 70 65 60 55 50 45 40 35 30

RAINFALL

Plants react to fluctuations in the temperature by shortening or lengthening their growing seasons. In warm years the latter are

by the method of least squares.

As a result of the appropriate comparisons we obtain the following regression equation:

$$y_{wt} = +0.094 - 0.04323(p_{VI} - 44) - 0.15276(p_{VII} - 57) - 0.08507(p_{VII} - 70) + 0.58602(t_{VII} - 17.6)(XII)$$

where y_{wt} is the index of productivity, p_{VI} , p_{VI} and p_{VII} are the rainfall for May, June and July respectively, and t_{VII} is the July temperature.

The correlation coefficient between y_{wt} (the index of productivity) and i is

$$R_{y, y_{wt}} = +0.62.$$

It is only slightly different from the correlation coefficient between yield and the aggregate of the weighted rainfall for the three-month period, which is 0.61.

This confirms the dominant, almost exclusive, influence of the rainfall during the period under study. The July temperature very slightly increases the closeness of the computed indices of productivity to the empirical, observed productivity.

Let us continue our investigation further. As has already been stated, the average date of the beginning of the rye harvest in Petrograd guberniya is 2 August, New Calendar. But the influence of the August weather on the rye yield is also observed, as may be seen from the following correlation coefficients:

August rainfall..... $r = +0.33$

August temperature..... $r = +0.34$

A negative effect of rainfall and a positive effect of temperature -- such is the action of the August weather.

This influence is partially explained by the fact that in years of late harvest rye matures during August, and partially by the fact that in ordinary years plants whose development has been retarded for one reason or another catch up and fully mature during the first days of August, but mainly by the fact that the period of harvesting and removal of rye from the fields falls during August. Rainfall and temperature are indices of how favorable the weather has been for this final sojourn of the rye on the fields. Dry, clear weather maintains the yield of rye. Bad weather and overcast skies threaten considerable crop damage from rotting and germination of the grain. Such phenomena are frequently observed in Petrograd guberniya. /1/

/1/ August is the rainiest month, in Petrograd guberniya. Cases of crop damage by unfavorable weather during the harvesting time and while the mowed grain is still in the fields occur, though considerably seldom, even in the dry chernozem belt. I recall 1896 and 1916, when considerable crop damage was caused by bad weather during this period. The investigation of the period of harvesting and removal of the grain requires the introduction of an additional factor in the shape of the number of rainy days occurring during that time.

The regression equation defining of the yield on the August rainfall and temperature, is as follows:

$$y_a = +0.029 - 0.0668(p_{VIII} - 78) + 1.46583(t_{VIII} - 15.1)$$

(XIV)

The correlation coefficient of the combined action of rainfall and temperature is:

$$r_{y, y_a} = 0.41.$$

We conclude this chapter by combining in a single table the factors that influence the yield of rye and giving their correlation coefficients (see Table 2)

T a b l e 2

FACTORS INFLUENCING THE YIELD OF RYE IN PETROGRAD GUBERNIYA

PERIODS AND FACTORS	Designation	r	R	REGRESSION EQUATION
Rainfall during fallow period	p_p	-0.46	0.52	$y_{pp} = -0.0919(p_p - 145).$
July temperature in the year preceding the harvest	$-1^{th} VII$	0.40		$y = -1^{th} VII = 1.718 * (-1^{th} VII - 17.7)$
Rainfall from beginning of final plowing for rye seeding, during sowing and to onset of tillering	P_{spg}	-0.57	0.74	$y_{spg} = 3.036 - 0.3995 (P_{spg} - 50) - 0.019034 (P_{spg} - 50)^2 + 0.00013861 (P_{spg} - 50)^3$
Rainfall of period of onset of tillering	$P \frac{3}{IX}$	0.34	0.34	$y = 1P_3 \frac{3}{IX} = +0.02 + 0.2307 (P_3 \frac{3}{IX} - 17)$
September and October temperature	$-1^{th} IX, X$	0.30	0.30	$y = -1^{th} IX, X = 2.010 * (t_{IX, X} - 8.9)$

FACTORS INFLUENCING THE YIELD OF RYE IN PETROGRAD GUBERNIYA

PERIODS AND FACTORS	Designation	r	R	REGRESSION EQUATION
---------------------	-------------	---	---	---------------------

Spring and Summer Period	May rainfall	P _V	-0.28	0.62	$y_s = 0.099 - 0.04323$
	June rainfall	P _{VI}	-0.50		$(p_{VI} - 44) - 0.15276$
	July rainfall	P _{VII}	-0.36		$(p_{VI} - 57) - 0.08509$
	July temperature	t _{VII}	0.26		$(p_{VII} - 70) + 0.58002$ $(t_{VIII} - 17.6)$

Absence of data on vegetative periods forces us to assume that dependence is reproduced weaker than reality.

Period of Harvest and Removal	August rainfall	P _{VIII}	-0.33	0.41	$y_a = + 0.029 - 0.0668$
	August temperature	t _{VIII}	0.34		$(p_{VIII} - 78) + 1.46508$
					$(t_{VIII} - 15.1)$

There are no exact data on limits of this period.

* In view of the absence of data on the vegetative periods, the dependence is reproduced more weakly than reality.

Let us now employ IV for the purpose of determining the expected yield, the dependences we have established.

The most exact method is to use the method of least squares to calculate the regression equations defining the yield as a function of all the meteorological elements indicated above. But the use of this method, with the large number of factors introduced, requires an immense amount of computation, justified only if there are no defects in the elements of observation.^{/1/} These defects are enumerated in Table 2 of the preceding chapter. I consider the most important of them to be the lack of information on the vegetative periods of the spring and summer life of winter crops. I have thus applied a simpler, approximate method in the present work. 1. I establish the indices for the four principal periods noted below, using the method of least squares. 2. From these indices I calculate a single composite index in accordance with the "weight" of the index for each period. 3. I connect the yield by a regression equation with this composite index.

^{/1/} 11 factors. In addition, in view of the curvilinear correlation for the period from the beginning of the last plowing for the sowing, down to the onset of tillering, it is necessary to bring in as special factors the second and third powers of the rainfall indices of this period.

We combine the meteorological elements into four principal periods.

I. The period before the beginning of the final plowing for

rye seeding. Meteorological factors: rainfall during the fallow period, temperature of the July of the year preceding the harvest. The regression equation defining the yield as a function of these factors is as follows:

$$IY = -0.122 - 0.07213(p - 145) + 1.4155 \\ (-1.2711 - 17.4)$$

The correlation coefficient that takes account of the combined action of these factors, i.e. the correlation coefficient between yield and the indices IY computed by the regression equation, is:

$$R_{-1} = 0.52.$$

II. The autumn period, from the beginning of the final plowing for rye seeding, to November. Meteorological elements: rainfall from beginning of final plowing for rye seeding to the onset of tillering; rainfall of the period of inception of tillering, and temperature for September and October.

The regression equation:

$$IIY = +0.222 + 0.93264Y_{spg} + 0.17705\left(\frac{p3}{IX} - 17\right) + \\ 0.3995(-1.2711, x - 8.9)$$

where Y_{spg} are the indices calculated from the regression equations 1/

1/ We have brought into the equation, as a factor, the indices of yield rather than the rainfall itself, in view of the curvilinear correlation. It would have been more accurate to bring in the first, second and third powers of the rainfall as special factors. But this would have unduly complicated the computation.

$\frac{3}{IX}$ is the rainfall of third September decade, and $t_{-1 IX, X}$ is the mean temperature for September and October. $R_2 = 0.79$.

III. The period of the spring and summer life of rye.
Meteorological factors: rainfall in May, June and July; temperature of July.

The regression equations:

$$III Y = +0.079 - 0.04333(P_{II} - 44) - 0.15276(P_{VI} - 57) - 0.08509(P_{VII} - 70) + 0.52002(T_{VII} - 17.6).$$

$$R_3 = 0.62$$

IV. The period of harvesting and removal of rye from the field. Meteorological factors: temperature and rainfall of August.

The regression equations:

$$IV Y = +0.024 - 0.0668(P_{VIII} - 78) + 1.46503(T_{VIII} - 15.1)$$

$$R_4 = 0.41$$

The yield indices Y_1 , Y_2 , Y_3 , and Y_4 , computed for each period, are given in Table 1, appended below. We may regard them as the most probable quantitative expression of the yield to be expected for the definite weather situation of a given period. But the degree of accuracy of the indices is different for each period. It is measured by the root-mean-square deviation of the observed yield from its "theoretical" norms, i.e. from the values of Y_1 , Y_2 , Y_3 , and Y_4 corresponding to each year; and is defined by the formula

$$\sigma = \sigma_y \sqrt{1 - R^2} \quad (\text{where } \sigma_y \text{ is the standard deviation of the values})$$

of y from their arithmetic mean). Consequently we have:

For the I period..... $\alpha_1 = \delta_y \sqrt{1 - 0.52^2}$

For the II period..... $\alpha_2 = \delta_y \sqrt{1 - 0.37^2}$

For the III period..... $\alpha_3 = \delta_y \sqrt{1 - 0.62^2}$

For the IV period..... $\alpha_4 = \delta_y \sqrt{1 - 0.41^2}$

By analogy to cases of an unchanging norm (arithmetic mean), we assume that the "weights" of the members of each series are inversely proportional to the squares of the measures of accuracy, i.e.

$$\mu_1 : \mu_2 : \mu_3 : \mu_4 = \frac{1}{\alpha_1^2} : \frac{1}{\alpha_2^2} : \frac{1}{\alpha_3^2} : \frac{1}{\alpha_4^2} = 1.37 : 2.66 : 1.62 : 1.20$$

The respective relative weights of the indices y_I, y_{II}, y_{III}

and y_{IV} are thus:

$\mu_1 = 1.37$, or, taking the sum of the weights as unity, to 0.20
 $\mu_2 = 2.66$, or, taking the sum of the weights as unity, to 0.39
 $\mu_3 = 1.62$, or, taking the sum of the weights as unity, to 0.24
 $\mu_4 = 1.20$, or, taking the sum of the weights as unity, to 0.17

Total= 6.85, or, taking the sum of the weights as unity, to 1.00

Using these "weights" we find a composite yield index for all four periods by summing the partial indices y_I, y_{II}, y_{III} and y_{IV} each with its corresponding weight $/1/$

$/1/$ This method is analogous to the derivation of a general mean value from partial means: if M_1, M_2 and M_3 , etc. are partial means, and $\alpha_1, \alpha_2, \alpha_3$, etc. are the standard deviations of the

partial means, then the weight $\mu_1, \mu_2, \mu_3, \text{etc.}$ of each mean is proportional to $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \text{etc.}$ while the general mean is defined by the formula

$$M_0 = \frac{M_1 \mu_1 + M_2 \mu_2 + M_3 \mu_3}{\mu_1 + \mu_2 + \mu_3}$$

i.e. $z = 0.20 \text{ I } y + 0.39 \text{ II } y + 0.24 \text{ III } y + 0.17 \text{ IV } y$

Thus, for example, for 1905, $\text{I } y = 6.0$; $\text{II } y = -15.8$; $\text{III } y = -2.5$; $\text{IV } y = -1.4$. Consequently the composite index z for 1905 is:

$$z = -0.20 \times 6.0 - 0.39 \times 15.8 - 0.24 \times 2.5 - 0.17 \times 1.4 \\ = -8.3$$

The composite index is given in column 7 of Table 1 of this chapter.

Finally we determine the regression equation connecting the yield with the composite index z :

$$y_z = 1.90 z.$$

The values of y_z , which constitute the quantitative expression of the yields to be expected, as determined by the totality of the action of all the meteorological factors that have been introduced into the investigation, are given in column 11 of the same table.

The correlation coefficient between y and z is:

$$r_{y,z} = 0.89.$$

Table 1

YEAR OF HARVEST	OBSERVED YIELD	I ^y	II ^y	III ^y	IV ^y	z	YEAR OF HARVEST	z	OBSERVED YIELD	y _z
1	2	3	4	5	6	7	8	9	10	11
1883	+ 0.1	+ 4.3	+ 1.4	- 2.1	- 2.1	+ 0.6	1895	- 8.6	- 15.3	- 16.4
1884	- 7.2	- 0.5	- 2.4	- 0.4	+ 0.4	- 1.0	1905	- 8.3	- 16.4	- 15.8
1885	- 1.5	- 1.6	- 6.0	+ 3.4	+ 1.8	- 1.5	1894	- 6.3	- 11.9	- 12.0
1886	+ 4.1	+ 6.1	+ 7.8	+ 2.6	+ 1.3	+ 5.0	1904	- 4.4	- 5.8	- 8.4
1887	+ 8.0	+ 0.7	+ 6.6	- 2.0	+ 0.2	+ 2.3	1907	- 4.2	- 4.3	- 8.0
1888	+ 8.4	- 2.8	+ 5.7	+ 5.0	+ 1.6	+ 3.1	1903	- 3.7	- 8.7	- 7.0
1889	+ 2.9	+ 0.9	+ 1.0	+ 6.6	- 0.9	+ 2.0	1906	- 2.8	- 4.8	- 5.3
1890	+ 10.2	+ 3.2	+ 4.1	+ 1.2	+ 3.6	+ 3.2	1893	- 2.1	- 5.6	- 4.0
1891	+ 8.7	+ 3.0	+ 5.5	+ 7.8	- 0.9	+ 4.4	1885	- 1.5	- 1.5	- 2.9
1892	- 4.1	+ 5.5	+ 4.6	- 8.5	- 1.9	+ 0.6	1897	- 1.4	+ 1.3	- 2.7
1893	- 5.6	- 7.2	+ 1.7	+ 0.4	- 0.2	- 2.1	1902	- 1.1	- 6.4	- 2.1
1894	- 11.9	+ 0.4	- 7.7	- 13.6	- 0.6	- 6.3	1908	- 1.1	- 2.1	- 2.1
1895	- 15.3	- 8.1	- 16.4	- 4.1	+ 2.5	- 8.6	1899	- 1.0	- 9.0	- 1.9
1896	+ 8.1	- 1.4	+ 2.3	+ 4.6	- 1.4	+ 1.5	1884	- 1.0	- 7.2	- 1.9
1897	+ 1.3	+ 4.3	- 5.6	- 2.1	+ 2.6	- 1.4	1913	- 0.2	- 1.1	- 0.4

Table 1

YEAR OF HARVEST	OBSERVED YIELD	I ^y	II ^y	III ^y	IV ^y	z	YEAR OF HARVEST	z	OBSERVED YIELD	y ₂
1	2	3	4	5	6	7	8	9	10	11
1898	+ 4.0	+ 0.9	+ 4.2	+ 0.9	+ 6.3	+ 3.1	1916	+ 0.4	- 3.3	+ 0.8
1899	- 9.0	- 0.2	- 1.5	- 0.1	- 1.8	- 1.0	1908	+ 0.5	+ 8.7	+ 1.0
1900	+ 8.1	+ 2.5	+ 4.9	+ 4.4	+ 4.5	+ 4.2	1892	+ 0.6	- 4.1	+ 1.1
1901	+ 7.7	+ 2.1	+ 5.9	+ 3.0	+ 3.4	+ 4.1	1883	+ 0.6	+ 0.1	+ 1.1
1902	- 6.4	+ 3.3	+ 2.3	- 8.1	- 4.2	- 1.1	1896	+ 1.5	+ 8.1	+ 2.9
1903	- 8.7	- 6.8	- 2.6	- 2.8	- 3.7	- 3.7	1912	+ 1.6	+ 6.6	+ 3.0
1904	- 5.8	- 5.0	- 3.5	- 4.0	- 6.6	- 4.4	1889	+ 2.0	+ 2.9	+ 3.8
1905	-16.4	- 6.0	-15.8	- 2.5	- 1.4	- 8.3	1910	+ 2.1	+ 1.7	+ 4.0
1906	- 4.8	- 2.1	- 1.1	- 5.3	- 4.1	- 2.8	1914	+ 2.2	+ 5.4	+ 4.2
1907	- 4.3	- 4.0	- 7.7	+ 1.7	- 4.0	- 4.2	1887	+ 2.3	+ 8.0	+ 4.4
1908	+ 8.7	+ 0.5	+ 2.9	- 2.6	- 1.0	+ 0.5	1911	+ 2.4	+ 2.6	+ 4.6
1909	- 2.1	- 2.6	- 2.9	+ 1.7	+ 1.0	- 1.1	1888	+ 3.1	+ 8.4	+ 5.9
1910	+ 1.7	- 0.6	- 4.9	+ 4.1	- 3.9	+ 2.1	1898	+ 3.1	+ 4.0	+ 5.9

Table 1

YEAR OF HARVEST	OBSERVED YIELD	I ^y	II ^y	III ^y	IV ^y	z	YEAR OF HARVEST	z	OBSERVED YIELD	y _z
1	2	3	4	5	6	7	8	9	10	11
1911	+ 2.6	+ 1.9	+ 4.4	- 1.4	+ 3.4	+ 2.4	1890	+ 3.2	+ 10.2	+ 6.1
1912	+ 6.6	- 4.1	+ 2.1	- 3.8	+ 4.2	+ 1.6	1901	+ 4.1	+ 7.7	+ 7.8
1913	- 1.1	+ 1.5	- 4.5	+ 1.8	+ 4.7	- 0.2	1900	+ 4.2	+ 8.1	+ 8.0
1914	+ 5.4	+ 2.5	+ 0.1	+ 6.9	- 0.3	+ 2.2	1891	+ 4.4	+ 8.7	+ 8.4
1915	+ 10.9	+ 7.4	+ 5.2	+ 2.3	+ 1.7	+ 4.4	1915	+ 4.4	+ 10.9	+ 8.4
1916	- 3.3	+ 2.0	+ 3.5	- 2.7	- 4.2	+ 0.4	1886	+ 5.0	+ 4.1	+ 9.5

In view of the uptrend in the rye yield in Petrograd guberniya, we operated with the deviations from the norms separately fixed for each year (see beginning of Chapter II). We pass now to the "absolute" yield in puds per desyatin, for which the expected yield, for each year, expressed in terms of deviations from the norm (see col. 11, table 1, this chapter), must be added to, or subtracted from, the normal yield for that year (see col. 6, table 1, chapter II). Thus, for example, the normal yield for 1905 is 49.1 puds, and the expected yield, as deviation from it, is 15.8 puds. The expected yield in puds per desyatin is consequently $49.1 - 15.8 = 33.3$. The normal yield for 1915 was 53.0, and the expected yield, as a deviation from it was 8.4. The expected yield in puds per desyatin is $53.0 + 8.4 = 61.4$. Table 2 below gives the indices of expected yield, thus computed, and compares them with the observed yield. The divergence between them, or residuals, are given in column 4 of the same table and denoted by the symbol alpha

(Table No 2, see following page)

Diagram 3 shows the observed yield (y series) and the "theoretical" yield (y_1 series)

All of the criteria establish the high degree of the parallelism between the observed yield and the quantitative expression of the expected yield determined on the basis of (1) the uptrend in the yield (see Table 1, Chapter II) and (2) the meteorological factors included in the investigation: the correlation coefficient between y_1 and y is expressed by the value 0.92, the root-mean-square deviation of the y series from the corresponding indices y_1 is equal to 3.32, and the mean deviation is 2.60 puds.

Table 2

YEAR	y	y ₁	a = y - y ₁	YEAR	y	y ₁	a = y - y ₁
1	2	3	4	1	2	3	4
1883....	40.4	41.4	-1.0	1900...	55.1	55.0	+ 0.1
1884....	33.5	38.8	-5.3	1901...	55.2	55.3	- 0.1
1885....	39.6	38.2	+1.4	1902...	42.5	45.8	- 4.3
1886....	45.6	51.0	-5.4	1903...	39.6	41.3	- 1.7
1887....	49.9	46.3	+3.6	1904...	42.9	40.3	+ 2.6
1888....	50.7	48.2	+2.5	1905...	32.7	33.3	- 0.6
1889....	45.6	46.5	-0.9	1906...	44.7	44.2	+ 0.5
1890....	53.3	49.2	+4.1	1907...	45.6	41.9	+ 3.7
1891....	52.2	51.9	+0.3	1908...	59.0	51.3	+ 7.7
1892....	39.8	45.0	-5.2	1909...	48.6	48.6	0.0
1893....	38.7	40.3	-1.6	1910...	52.8	55.1	- 2.3
1894....	32.8	32.7	+0.1	1911...	54.1	56.1	- 2.0
1895....	29.8	28.7	+1.1	1912...	58.4	54.8	+ 3.6
1896....	53.6	48.4	+5.2	1913...	51.1	51.8	- 0.7
1897....	47.2	43.2	+4.0	1914...	58.0	56.8	+ 1.2
1898....	50.3	52.2	-1.9	1915...	63.9	61.4	+ 2.5
1899....	37.7	44.8	-7.1	1916...	50.1	54.2	- 4.1

y : series of observed yield;

y₁: theoretical yield established on basis of the uptrend in the yield of Petrograd guberniya and of the meteorological factors included in the investigation.

alpha: divergence between the series y and y₁ ; $\sum a^2 = 374$; $\sigma_a = \sqrt{\frac{374}{34}} = \sqrt{11} = 3.32$

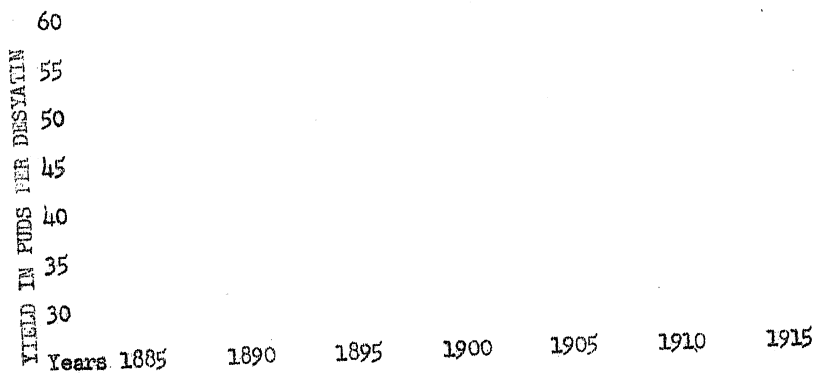
$\sum [x] = 88.4$
Mean deviation: $\bar{y} = \frac{\sum [x]}{34} = \frac{88.4}{34} = 2.60$

Coefficient of correlation between y and y₁ = 0.92

D i a g r a m 3

YIELD OF RYE IN PETROGRAD GUBERNIYA FROM DIRECT OBSERVATIONS AND AS
CALCULATED ON THE BASIS OF THE METEOROLOGICAL FACTORS

----- Observed yield
____ Yield calculated on the basis of the meteorological factors



In some cases, however, the difference between the expected and observed yield reaches a fairly high value: in two cases it is 7.1 and 7.7 puds, and in four /other/ cases it is over 5 puds (5.2 - 5.4 puds). This compels us to postulate the presence of certain substantial factors not included in the investigation. And, in fact, lacking factual material on the subject, I had to leave out the winter period. Yet under the climatic conditions of Petrograd guberniya this period plays a great role in the life of winter crops. Snowfall on thawing land and the formation of an ice-crust as a result of winter thaws, or of the slow thawing of the snow in the spring, lead to the death of winter crops by suffocation /VYPREVANIYE/ or insufficient access of air. Winterkilling, through freezing, occurs far less frequently in Petrograd guberniya and

does far less damage to winter crops there. Unfortunately I lack empirical material to characterize the winter damage throughout the entire 34-year period. Only since the organization of the current statistical service of the Zemstvo in Petrograd guberniya, in the agricultural year 1894/95, has it been possible to obtain, if not entirely systematic information on the winter period of winter crops, at any rate information sufficiently complete. But even this information is not expressed in the form of numerical indices, and these must be extracted from the text descriptions of the conditions of growth of the winter crops /1/.

/1/ In 1916 the current statistical service of the zemstvo in Petrograd guberniya did not function (they only collected crop data). I was forced to fill in the gap by personally questioning a number of local inhabitants about the winter damage

During the 22 year period from 1895-1916, the rye crops suffered badly in the winters of 1902 and 1899; appreciable damage was observed in 1909, 1910, 1913, and 1916. For the remaining 12 years I did not find definite indications of winter suffocation in the statistical summaries of /VYPREMANIYE/, winterkilling through freezing, or of the soaking /VYMOCHEKA/ of winter crops at their emergence from the snow. It must be assumed that the wintering conditions in these years were favorable. If we denote by the symbol b the winter damage to winter crops, and we give the value b a numerical value of minus 2 in cases of serious damage, minus one in cases of appreciable damage and zero for years with favorable wintering conditions for rye sowings. We determine the

correlation coefficient between y (the yield expressed in deviations from normal) and b .

$$R_{y,b} = +0.29$$

Thus, the more favorable the wintering conditions are, the better the yield.

We find the regression equation that define the yield as a function, firstly of the yield indices set up on the basis of the dependence of the rye harvest on the meteorological factors included in the investigation (see series y of table 1 of this chapter), and secondly of the values b :

$$y_b = +1.66 + y_z + 3.6223b$$

The values of y_b are given in table 3, in which also their divergencies from the observed yield (series alpha) are also shown. The mean deviation between the theoretical and the observed yields is $\alpha_{\text{sub zero}} = 2.05$, and the standard deviation 2.42 for the same two years, the theoretical values for the yield, determined on the basis of the course of the meteorological elements without the winter period, diverge on the average from the empirical yields by 2.59 (mean deviation) and 3.33 (standard deviation).

Thus the inclusion of the winter period, as should have been expected, increases the accuracy of the investigation.

Table 3

YEAR	y_b	y	$y - y_b = a$	YIELD IN PUDS PER DESYATIN		YEAR	y_b	y	$y - y_b = a$	YIELD IN PUDS PER DESYATIN	
				y_2	y					y_2	y
				5	6					5	6
1	2	3	4	5	6	1	2	3	4	5	6
1895	- 14.8	- 15.3	-0.5	30.3	29.8	1906	- 3.6	-4.8	- 1.2	45.9	44.7
1896	+ 4.5	+ 8.1	+3.6	50.5	53.6	1907	- 6.3	- 4.3	+ 2.0	43.6	45.6
1897	- 1.1	+ 1.3	+2.4	44.8	47.2	1908	+ 2.7	+ 8.7	+ 6.0	53.0	59.0
1898	+ 7.5	+ 4.0	-3.5	53.8	50.3	1909	- 4.1	- 2.1	+ 2.0	46.6	48.6
1899	- 7.5	- 9.0	-1.5	39.2	37.7	1910	+ 2.0	+ 1.7	- 0.3	53.1	52.8
1900	+ 9.6	+ 8.0	-1.6	56.7	55.1	1911	+ 6.3	+ 2.6	- 3.7	57.8	54.1
1901	+ 9.5	+ 7.7	-1.8	57.0	55.2	1912	+ 4.7	+ 6.6	+ 1.9	56.5	58.4
1902	- 7.7	- 6.4	+1.3	40.2	41.5	1913	- 2.4	- 1.1	+ 1.3	49.8	51.1
1903	- 5.3	- 8.7	-3.4	43.0	39.6	1914	+ 5.9	+ 5.4	- 0.5	58.5	58.0
1904	- 6.8	- 5.8	+1.0	41.9	42.9	1915	+10.1	+10.9	+ 0.8	63.1	63.9
1905	- 14.1	- 16.4	-2.3	35.0	32.7	1916	- 1.2	- 3.3	- 2.1	52.2	50.1

y : is the observed yield expressed as deviations from normal

y_2 is the theoretical yield.

There can be no doubt that eliminating the defects of observation, transition from approximate methods to more exact ones, and inclusion of new factors should increase even more the coincidence of the "theoretical yield" with the observed yield/1/

/1/ I consider the assignment of the meteorological factors to definitely established vegetative periods of the spring and summer life of plants to be extremely essential for the increasing of the accuracy of the regression equation. I have not been able to do this in the present work because I did not have the factual material. It is also possible that under these conditions a certain curvilinear nearity of the correlation for the spring and summer rainfall and for the influence of temperature of this period will be found, which has hitherto escaped notice under the present state of the factual material. It is just as important to supplement the investigation by information on the winter conditions (particularly on the snow-suffocation /Vyprevaniye/ of the winter sowings), which it was possible to do (and that only in part) during the 22-year interval. The analysis of the effect of November and winter rainfall would also not be superfluous. Moreover as we shall see below, the question as to the influence of the yield of prior years also arises. Finally it is very important to give a parallel variant for the combined influence of rainfall and temperature. In the present work, their influence has been considered apart from one another, and there was an entirely valid reason, in itself, for doing this, since the effects of rainfall and temperature are very different from each other in Petrograd guberniya. But the influence of rainfall is intensified or attenuated according to the temperature (and also according to the influence

of atmospheric humidity and of wind force). For this reason a second variant of the rainfall effect should be given, with corrections for the temperature (and it proves possible, with corrections for atmospheric humidity and wind force as well).

In conclusion, I present a brief determination of the expected yield on the basis of the course of the meteorological factors and of the correspondence estimates of crop condition. This could be done for a 14 year period for Petrograd guberniya. In table 4, series X gives the indices of yield based on the course of the meteorological elements up to the onset of winter weather (the fallowing period and the period of autumn weather), the series D shows the indices of yield, computed on the basis of autumn and summer weather (the period of spring and summer life of winter rye, and the period of the harvest.) These indices have been established by exactly the same method as the general index (see tables 1 and 2 of this chapter). Series U gives the index of the expected yield on the basis of the correspondence estimates of crop condition after emergence from the snow /1/.

/1/ See chapter I, table 6, column 3

On connecting these indices by regression equation according to the method of least squares, we have

$$y_3 = 49.5 + 0.5903(x - 47.5) + 0.5600(z - 47.5) + 0.3769(y - 49.5)$$

The divergence between the theoretical and observed indices of yields were subjected by me to further statistical analysis, in the course of which I apparently succeeded in noting the negative effect on the rye yield exerted by the preceding crop rotation. The correlation coefficient between the value of U and the yield of rye in the preceding crop rotation is $-0.42 / 2/$.

2/ The yield of rye in 1902 was compared with the yield of 1899, the yield of 1903 with the yield of 1900, the yield of 1901 with the yield of 1901, etc. The yield of oats does not give such an effect (a comparison was made of the yield of rye in 1902 with the yields of oats in 1900, of the yield of rye in 1903 with the yield of oats 1901). It is possible that a good yield of oats, even though it depletes the fertility of the soil, still yields more manure from the use of the straw and grain on the farm when the crop is good, and that this manure then goes to fertilize the same field from which the oat crop was taken. With a poor oat crop, even though less fertilizer is obtained, the soil is less exhausted. Thus the exhaustion of the soil and its fertilizing are mutually compensated.

It is obvious that the investigation requires supplementary work with respect not only to the yields of the preceding crop turn-over but also with respect to fertilizer. The yield of rye straw and hay in the year preceding the fallowing of the field affects the amount of fertilizer on the fallow field to be sown to rye. For this reason it would be curious to trace the influence of the harvest of rye straw and hay in 1900 on the rye yield of 1902, the harvest of rye straw and hay of 1901 on the rye yield of 1903, etc.

Consequently after a poor rye crop in the preceding crop rotation, we have a certain favorable chance for a good yield of rye in the following crop rotation, and conversely, a good yield of rye creates unfavorable chances for the yield of rye in the following crop rotation. If this is confirmed by further investigation, then, we are obviously dealing in this case with an effect of good yields exhausting the soil, and an action of poor yields to spare it. The indices of yield, corrected for the effect of the rye yield in the preceding crop rotation are given in column 8 of table 4. Column 9 shows the divergences of these indices from the observed yields. Column 10 gives these divergences in per cent of the theoretical yield. As we can see, the largest divergences do not reach 10 per cent (minus 8.0 and plus 7.0) and their mean value is 3.5 per cent.

With this I conclude this investigation and allow myself to affirm that the application of the methods of mathematical statistics has given very satisfactory results. There used on an even wider scale and more complete material will yield even more fruitful result.

Originally published in Vestnik St VESTNIK STATISTIKA, No. 1-4, 1921.

OF THE USE OF SAMPLING METHODS

We have a population consisting of m units. A certain phenomenon is registered by a set of instruments covering the entire field, which however are defective and introduce a systematic error into the results of the investigation. Assume that in order to ascertain and determine the extent of the systematic error in this mass coverage, which is defective, a supplementary examination is organized, based on a random selection (for instance, by drawing lots) of n units out of the m elements of the population. This sampling is based on a precise method eliminating the systematic error (for instance, by instrumental survey of seeded acreage). Thus we have two parallel series of observations, y and x , the series y from an exact source, which does not, however, go beyond the limits of n sampling observations, and the series x , with observations from a source giving the same n objects full but defective coverage.

$$\begin{array}{l} x_1 \text{ --- } y_1 \\ x_2 \text{ --- } y_2 \\ x_3 \text{ --- } y_3 \\ \vdots \\ x_n \text{ --- } y_n \end{array}$$

The summarizing criteria of these series:

Arithmetic means: $M_x \dots\dots M_y$

Standard deviations: $\sigma_x \text{ --- } \sigma_y$

OF THE USE OF SAMPLING METHODS

We have a population consisting of m units. A certain phenomenon is registered by a set of instruments covering the entire field, which however are defective and introduce a systematic error into the results of the investigation. Assume that in order to ascertain and determine the extent of the systematic error in this mass coverage, which is defective, a supplementary examination is organized, based on a random selection (for instance, by drawing lots) of n units out of the m elements of the population. This sampling is based on a precise method eliminating the systematic error (for instance, by instrumental survey of seeded acreage). Thus we have two parallel series of observations, y and x , the series y from an exact source, which does not, however, go beyond the limits of n sampling observations, and the series x , with observations from a source giving the same n objects full but defective coverage.

$$\begin{array}{l} x_1 \text{ --- } y_1 \\ x_2 \text{ --- } y_2 \\ x_3 \text{ --- } y_3 \\ \vdots \\ x_n \text{ --- } y_n \end{array}$$

The summarizing criteria of these series:

Arithmetic means: $M_x \dots M_y$

Standard deviations: $\sigma_x \dots \sigma_y$

Standard deviations of the arithmetic means:

$$\bar{E}_x = \frac{\sigma_x}{\sqrt{n}} \dots \bar{E}_y = \frac{\sigma_y}{\sqrt{n}}$$

Let x_0 be the mean value of the total population of the series

x.

Then, on the basis of: (a) this completely covered but defective source, and (b) a correction established from the sample, but precise, examination, we may determine the measurements of this population which are closest to reality, namely:

$$A_x = m(M_y - M_x + x_0),$$

where m is the number of units in the population. But we may proceed to determine the value of A in a different way: rejecting the defective mass-coverage source x , we may make a sampling investigation of the phenomenon only with a precise set of instruments, and then multiply the mean value of the sample M_y by the number of units in the population:

$$A_y = mM_y.$$

The number of units in the population is the same, and the non-coincidence of the results A_x and A_y depends on the multipliers: $M_y - M_x + x_0$ in the former case and M_y in the latter.

Which of these two methods should be used to determine these

measurements of the phenomenon under study which most closely correspond to reality?

Obviously the method giving more accurate results for the same size of sample is preferable.

Mathematically this condition is expressed as follows:

$$\text{If: } E(m_y - m_x) = \sqrt{E_x^2 + E_y^2} - 2E_y E_x r_{xy} > E_y$$

(1^a)

then the method of corrections should be rejected, and the examination of the sample, carried out with precise instruments, should be used as an independent source.

$$E(m_y - m_x) = E_y$$

(1^b)

the two methods are of equal value.

Finally, if

$$E(m_y - m_x) < E_y$$

(1^c)

Then the correction method has the advantage, and the examination of the sample by precise instruments, is used only as a control.

By a simple algebraic transformation of the expressions (1) we obtain the following conditions under which one variant or the other is possible:

$$r_{xy} < \frac{1}{2} \frac{E_x}{E_y}$$

(2^a) the advantage is on the side of independent computation by the sampling method

$$r_{xy} = \frac{1}{2} \frac{E_x}{E_y}$$

(2^b) the two methods are of equal value

$$r_{xy} > \frac{1}{2} \frac{E_x}{E_y}$$

(2^c) the advantage is on the side of the correction method

(In organizing work with precision instruments it is in practice impossible to take samples amounting to a very high percentage of the population, owing to limitations funds and cadres of qualified workers' taken as so that the standar deviations of the means M_y and M_x may be/equal to

$$\frac{\sigma_x}{\sqrt{n}} \quad \text{and} \quad \frac{\sigma_y}{\sqrt{n}}$$

Usually σ_x and σ_y , and consequently also E_x and E_y (the number of observations in the x and y series being equal) differ very little in practice, since the series x and y both consist of observations on the same criteria of the same objects, while the defectiveness in the

methods of mass observation lead, where tolerated, to a systematic error but not to a more considerable random error. Let us therefore first analyze the case where $E_x = E_y$. Here the preceding expressions are simplified, and we have, respectively:

$$r_{xy} < 0.50$$

(3^a)

$$r_{xy} = 0.50$$

(3^b)

$$r_{xy} > 0.50$$

(3^c)

Thus the superiority of the one method or the other depends on the extent to which parallelism is observed between the defective series x and the series y. If this parallelism is weak and is expressed by a coefficient of correlation less than +0.50, the employment of the

coefficient of correction should be rejected; with $r_{xy} = 0.50$ comes the changeover, at which the two methods are of equal value; and as the parallelism between the series x and y increases further, the advantage passes more and more to the correction method.

In general, by analyzing the expression $E(m_y - m_x)$, which, when $\sigma_x = \sigma_y$, is transformed to the formula

$$E(M_y - M_x) = E_y \sqrt{2(1 - r_{xy})}$$

(4)

the limiting values of the advantage of one method or the other may readily be established.

The correlation coefficient increases algebraically from - 1.00 to + 1.00. The value of $E(m_y - m_x)$, as can be seen from formula (4), reaches its maximum for $r_{xy} = (-) 1.00$. In this case $E(m_y - m_x) = 2E_y$.

Thus, in the limiting case, the accuracy of calculation by the method of corrections proves to be half that of independent calculation from the examination of the sample, if the systematic error has been elimination.

It is, however, entirely clear that a negative correlation (to say nothing of one equal to unity) between the defective observations and the precise material would in fact be a completely impossible phenomenon. This would imply that the defective observations yielded not only an inaccurate but also an inverse, "negative" reflection of

reality. If the n observations of the precise sample examination are arranged in the order of their increasing value, so the $y_1 < y_2 < y_3 \dots < y_n$ then if $r_{xy} = (-)1.00$, the corresponding observations of the defective source would be arranged in the reverse order: $x_1 > x_2 > x_3 \dots > x_n$ and with r_{xy} lying within the limits of -1.00 and 0 , there would be a similar order of inequalities, though only approximately, and with exceptions. The presence of a negative r_{xy} would signify that the defective examination gave smaller indices where in reality the phenomenon with which we are concerned was quantitatively greater, and conversely.

Thus, if $\sigma_x = \sigma_y$, the method of corrections can never be twice as good as direct calculation from the sample examination, or, differently expressed, the results of independent calculation from the sampling instruments can never be twice as accurate as those obtained by using the sample observations as a control to establish coefficients of correction to the defective continuous-coverage examination.

Continuing our analysis of formula (4), we note that $E(m_y - m_x)$ declines as r_{xy} increases algebraically from -1.00 , and that, at the moment r_{xy} crosses zero, $E(m_y - m_x) = E_y \sqrt{2}$

$$\text{or } E_y = \frac{E(m_y - m_x)}{\sqrt{2}} = \frac{E(m_y - m_x)}{1.414}$$

However even the case $r_{xy} = 0$ must be considered improbable, since this would be evidence that the defectiveness of the instrumentation of the full-coverage study was so great that it was not in a position to reflect, to any extent whatsoever, the degree of actual variation of the criterion being studied. With ascending values of the precise series

the values of x would not even show a tendency to increase -- this is the quality of the full-coverage study that would be reflected by a r_{xy} equal to zero [If the series y be arranged in ascending order and then divided into s groups, and the mean of each group be determined, then $M_1 < M_2 < M_3 < \dots < M_s$ for all values of s . The corresponding means for x : $M_{x1}, M_{x2}, \dots, M_{xs}$ with $r_{xy} = 0$, approach approximate equality as s decreases and as the number of observations in each group correspondingly increases. In general, with $r_{xy} = 0$, we have before us the "ideal" case of statistical investigation "from ceiling".]

We know that the autumn questionnaires used for a number of years by the Central Statistical Administration gave yield indices which were extremely low by comparison with the actual yields; we also know that the estimates of seeded acreage made by the Marmkomfin often left much to be desired with respect to quality. But neither of them ever reached that stage of "laxity" at which they would not be in any position at all to reflect in their indices the differences of geography or of farms.

Our analysis thus leads us to the following conclusion: the limiting relative accuracy [The author measures accuracy by the value of $\frac{1}{E}$] (Editor's note). of direct calculation from a sample, study but an accurate one, as compared with the method of corrections, is expressed by a coefficient lower than $\sqrt{2} = 1.414$ (when $\sigma_x = \sigma_y$).

Let us see what the possible limits are for the superiority of the correction method over the method of computation from the sampling examination. The ratio $\frac{E(m_y - m_x)}{E_y}$ is a criterion for preferring one method or the other. If the mean error of the differences of two arithmetic means calculated for the same value of n or number of

observations of the precise sample and the mass-coverage study is greater than the mean error of the arithmetic mean found by precise examination, then

$$E(m_y - m_x) > 1$$

and in this case, as has been repeatedly pointed out, independent computation from the sampling data should be employed. If the opposite is observed and

$$E(m_y - m_x) < E_y$$

while the ratio $\frac{E(m_y - m_x)}{E_y} < 1$

then preference should be given to the method of corrections, and the precise sampling should be used as a control.

From Formula (4) we find

$$K = \frac{E(m_y - m_x)}{E_y} = \sqrt{2} \sqrt{1 - r_{xy}} \quad (5)$$

The right side shows that the criterion is exclusively determined by the size of the correlation coefficient between x and y. By using formula (5), we easily find from it all the partial cases considered above.

When: $r_{xy} = -1.00$ $K = 2$
 $r_{xy} = 0$ $K = \sqrt{2}$
 $r_{xy} = 0.50$ $K = 1$
 $r_{xy} > 0.50$ $K < 1$

Formula (5) makes it possible to determine the value of the criterion K for different values of r_{xy} . It is more convenient, however, for cases when $r_{xy} > 0.50$, to make use of the inverse value: $\frac{1}{K} = \frac{E_y}{E(n_y - n_x)}$, which shows by how much the method of independent computation from the sampling is less accurate than, and consequently inferior to, the method of corrections.

Here is an illustrative table of the values of $\frac{1}{K}$ for various values of r_{xy} :

r_{xy}	$\frac{1}{K^2}$	$\frac{1}{K}$
0.50	1.000	1.000
0.60	1.250	1.118
0.65	1.428	1.195
0.70	1.667	1.291
0.75	2.000	1.414
0.80	2.500	1.581
0.85	3.333	1.826
0.875	4.000	2.000
0.90	5.000	2.236
0.95	10.000	3.162
0.96	12.500	3.536

0.97	16.667	4.002
0.98	25.000	5.000
0.99	50.000	7.072
1.00	infinity	infinity

As may be seen from this table, the relative superiority of the method of corrections increases strongly with the correlation coefficient between the series x and y . The number of observations necessary to attain the degree of accuracy of the correction method from independent computation by the sampling method increases to an even greater degree. Thus, for example, if $r_{xy} = 0.75$ (i.e. $\frac{1}{\sqrt{2}}$) then the standard deviation of the independent computations is double that of the correction method. To attain the same result by direct computation by the sampling method, the size of the sample would have to be not double but fourfold, since the measure of accuracy is proportional to the square root of the number of observations. The value of $\frac{1}{\sqrt{2}}$ also shows how many times the number of observations in independent use of sampling must be increased to obtain the same results as by the method of corrections. [These attempts to catch up with the method of corrections are completely hopeless, however, in cases where the mass observations, though imperfect, are still necessary in themselves. Thus, for instance, the registration by the Narkomfia of acreage seeded for each farm is necessary for tax purposes, independently of whether or not they will be arranged in any particular number of seeded acreage measurements. For this reason increasing the number of objects measured, for instance, from n to $4n$ (with $r_{xy} = 0.875$), with the object of catching up with the method of errors, would inevitably involve using that very method for its own purposes, for the corresponding readings in the mass observation may be found for all of the additional $3n$ objects.]

In the last analysis the accuracy of the results of the methods of independent computation from sampling and of corrections increase to the same extent, and thus the former method will never succeed in

catching up with the latter.

A somewhat different situation arises with a defective continuous coverage (or an "intensified" mass coverage), the special-purpose set-up of which is closely connected with a supplementary precision sampling. In this case the renunciation of mass registration results in a saving of money and labor, which can be used to increase the scale of the precision sampling. Whether mass studies should be renounced in favor of broadening the scope of the sampling or not depends on whether it is possible with the funds thereby released to expand the sampling sufficiently to make the new computation (now already an independent one) yield more accurate results than the method of corrections with the former smaller sample. The decision here depends (1) on the value of the correlation coefficient between the defective observations (series x) and the accurate ones (series y): the higher this coefficient is, the less the chances are for a renunciation of mass observation, for the greater the number of supplementary objects that the precision sampling must cover to "overtake and surpass" the method of corrections; (2) on the relative cost of the more complete and detailed sampling study, which, when the object of the investigation is taken into consideration, is considerably more expensive than the solid-coverage (than the "strongly" mass study) by imperfect instrumentation.

Let B be the cost of the whole mass study, and $(\frac{1}{k^2} - 1)n$

the number of supplementary objects of the sampling necessary, with the corresponding value of r_{xy} , to attain the required accuracy of the results of computation by the correction method with a smaller number of observations. It is clear that renunciation of mass registration and

adoption of independent computation based on an increased volume of sampling will be advantageous if the cost of the study is less than

$$\frac{B}{(\frac{1}{K_2} - 1)n}$$

end of footnote]

If $r_{xy} = 1$, then $K = E(m_y - m_x) = 0$.

In this case the error of difference between the two means M_y and M_x is zero and consequently the method of corrections is absolutely accurate. It will be shown in the following chapter, first, that this absolute accuracy of the method of correction is conceivable only if $\sigma_x = \sigma_y$ if σ_x and σ_y are not equal, then even if $r_{xy} = 1$, the limiting minimum value of $E(m_y - m_x)$ cannot drop to zero, and that, secondly, in case $\sigma_x = \sigma_y$, a correlation coefficient equal to unity is possible only when the x series coincides with the y series, i.e. in essence is not defective, or if the defectiveness is exceedingly peculiar, namely in that all members of the series x shall differ from the series y by the same absolute value, i.e. $x_1 = y_1 + a$; $x_2 = y_2 + a$; etc.; $x_n = y_n + a$.

In real life such a rigid and petrified defectiveness is never met. Thus the advantage of the method of corrections, expressed in mathematical language, can never become infinitely great (with $E(m_y - m_x) = 0$, $\frac{1}{K} = \infty$), but it may still be expressed by very large numbers, as may be seen from the table of I and $\frac{1}{K}$ presented above; everything depends on the value of r_{xy} . We recall that the limits of the advantage of the independent computation from precise sampling are incomparably narrower: its accuracy as compared with the method of corrections can never at-

tain $\sqrt{2}$ (even if a negative r_{xy} were considered possible, down to a value of -1 for r_{xy} , even then it still could not exceed 2). But a numerical expression, as large as this, for the advantage of the method of corrections, comes into force already with $r_{xy} = 0.750$. But it may perhaps that so "high" a degree of parallelism between the readings of the same criterion, on the same object, by the defective mass observation, on the one hand, and the precision instruments, on the other hand, could never be realized. Such a position would not hold against the lightest touch of criticism, for it is equivalent to the affirmation that in reality only such defectiveness of mass investigation is possible as would be responsible for a correlation coefficient of less than 0.75. (To anticipate somewhat, we may note that even so "high" a correlation coefficient as 0.75 would still evidence a very great defectiveness in the continuous-coverage observation, since it would indicate that even under the most favorable conditions reality would be reflected in its readings only to the extent of 56 percent, while 44 percent represents distortions.) between the series and would exclude all degrees of defectiveness allowing higher correlation coefficients than that.

Work-up of the surveys of seeded acreage in 1930, made in Belorussia, Ukraine and the Middle Volga, gives the following correlation coefficients between the surveys and the area registered by the Narkomfin (The comparison is arranged by the village soviets in which the surveys of seeded acreage were made. The mean indices of sowings per farm were calculated for each village soviet from the tax rolls and the surveys, and these two series then correlated.)

Narkomfin:

Belorussia	0.897
Ukraine (as a whole)	0.986

including:

Polesia	0.912
Right Bank of Volga	0.952
Left Bank of Volga	0.985
Steppe	0.976
Middle Volga	0.971

As may be seen from this table, all these correlation coefficients decisively exceed 0.75, while in three cases (The Ukraine as a whole not considered in this connection.) they even exceed 0.97: the method of corrections has an immense advantage over the independent calculation of the size of the seeded areas from sample surveys (It must not be thought that the Narkomfin registrations of sown acreage in these regions are superlative and need no corrections; on the contrary, they are far from being perfect. The coefficients of acreage missed resulting from the surveys are as follows: Belorussia, 1.133; The Ukraine (all) 1.050, including: Polesia, 1.111; Right Bank of Volga 1.046; Left Bank of Volga, 1.035; Steppe, 1.050; Middle Volga, 1.051.).

We conclude this chapter with a short resume of its contents:

1. The method of independent computation is advantageous where $r_{xy} < 0.50$, while its limits of accuracy, as compared to the method of corrections, are very limited and do not attain

$\sqrt{2} = 1.414$, while its "coefficient of useful action" (meaning by this term the coefficient of increase in the size of the sample required to obtain the same degree of accuracy as by the method of corrections) can never exceed 2.

2. When $r_{xy} > 0.50$, the advantage passes over to the method of corrections. This advantage increases with r_{xy} ; with $r_{xy} = 0.75$, it reaches a value equal to the limiting accuracy of independent computation, and subsequently, as r_{xy} further increases, the "coefficient of useful action" for the method of corrections is expressed by very high figures (See table of values for $\frac{1}{k^2}$).

3. High correlation coefficients that unconditionally give the primacy to the method of corrections are not only theoretically possible, but are also frequently met in practice.

4. It is apparent from all that has been set forth that in general (but only in general) the method of corrections outweighs that of independent computation from precision sampling. This conclusion has not only formal character: it is founded on the undisputed fact that even the most imperfect full-coverage survey reflects reality, though in distorted form, thereby creating elements of identity between the indices of a defective investigation and those of a precision sampling. Measured by the community thus created, correlation likewise arises between the readings of the two sources.

5. All that has been set forth in this chapter holds only for $\sigma_x^2 = \sigma_y^2$.

What modification is introduced when this equality is violated?

The second chapter gives the answer to this question.

II

Let us return now to formula (2^b), which expresses the condition that the two methods shall be of equal value for use: $r_{xy} = 1 - \frac{E_x}{E_y}$. It is clear from this that, if $E_x = E_y$ (or, what is the same thing, if $\sigma_x = \sigma_y$, the changeover in favor of the method of corrections begins at higher correlation coefficients. Thus if the variability of the defective series x exceeds that of the precise series y , then the advantage of the method of correction is attenuated, and the more so, the higher the ratio $\frac{E_x}{E_y}$. (The ratio $\frac{E_x}{E_y}$ allows comparison between the accuracy of observation of continuous coverage and sampling studies to be made, assuming that the methods of observation were different in the two cases. E^2 depends on σ^2 , while the latter is made up of the real dispersion of the criterion σ_r^2 and the error of observation σ_o^2 . The latter values are different for continuous coverage studies and sampling studies.)

(Editorial Note) (Russian)

The following schematic table gives the numerical values of the coefficients of correlation at which this changeover occurs:

$\frac{E_x}{E_y} = 1$	r_{xy}
1.0	0.50
1.1	0.55
1.2	0.60
1.3	0.65
1.4	0.70
1.5	0.75
2.0	1.00

It is an interesting case when the variability of the imperfect mass observation is more than twice σ_y . With this ratio of the variabilities, the advantage always remains on the side of the independent computation based on precision sampling, since its relative advantage breaks off for $r_{xy} = 1$.

Such closeness of connection between the readings of the defective and the accurate studies does not occur in practice, though, as we have seen, the correlation coefficients between x and y may be very close to unity. It is likewise entirely clear that if $L > 2$, then a fortiori, there is no reason for using the method of corrections, since the changeover can appear only at the impossible numerical values of $r_{xy} > 1$.

It is apparent from the same formula $r_{xy} = 1 - \frac{E_x}{2} \frac{E_y}{E_x}$ that if the variability of the x series is less than that of the y series, ($\sigma_x < \sigma_y$, and consequently $E_x < E_y$) then inversion in favor of the method of corrections takes place with correlation coefficients less than 0.50, and its advantage in this connection is inversely proportional to the value of L .

$\frac{E_x}{E_y}$	r_{xy}
0.9	0.45
0.8	0.40
0.7	0.35
0.6	0.30
0.5	0.25

But in another respect its deteriorates. As was remarked with $\sigma_x = \sigma_y$, the error of difference of the means M_x and M_y

approaches zero as r_{xy} approaches 1, and since coefficients r_{xy} very close to 1 are not infrequently encountered, it is impossible to define in advance the limits of accuracy of the computation by the method of corrections; it may be exceedingly high. In cases where σ_x is not equal to σ_y , the limits of the maximum advantage of the method of correction is expressed by a finite number, namely: the error of the difference of the means M_x and M_y approaches the difference of the errors of the means and as r_{xy} approaches 1:

$$E(m_y - m_x) = E_y - E_x \quad (r_{xy} = 1)$$

(1)

It is easy to convince oneself of this from the formula:

$$E(m_y - m_x) = \sqrt{E_y^2 + E_x^2 - 2E_x E_y r_{xy}}$$

(2)

Putting $\frac{E_x}{E_y} = L$, we have:

$$K = \frac{E(m_y - m_x)}{E_y} = \sqrt{L^2 + 1 - 2Lr_{xy}} \quad (6)$$

We now have before us a formula defining in a general form the ratio of the errors of calculation by the method of corrections and by the method of independent computation. This ratio is a function of two variables: (1) the correlation coefficient between the readings from the source x and y ; and (2) their relative variation [or coefficient of variation]. If $L = 1$, we have the case considered in Chapter I and mathematically formulated in Formula (5).

If $r_{xy} = 1$, we have:

$$K = \frac{E(m_y - m_x)}{E_y} = \pm (1 - L) \quad (7)$$

(If $L > 1$, the sign is -; if $L < 1$, the sign is +)

All cases of limiting ratios, including also those previously discussed, are envisaged by this expression: if (1) $L = 1$, then $K = 2$; if $L > 1$, then if $L = 2$, even at the "limit", there is no more place for the advantage of the method of corrections; if, however, $2 < L < 1$, then K , appearing as a limiting value, is expressed as a fraction, and thus the limiting value of K marks the superiority of the method of corrections; finally, (3) if $L < 1$, then K is likewise a limiting and fractional number.

It should be noted that the more L differs from unity, whether above it or below it, the more strongly grows the value of K .

Thus, alongside of the disturbance of the equality of σ_x and σ_y (and as this disturbance increases), the "limiting" advantage of the method of corrections declines. Let $L_1 = 1 + \kappa$, and

$L = 1 - \dots$, where $0 \dots 1$. In these cases the extreme values of K are equal to the same value of $\dots$, as is evident from Formula (7). Thus, for instance, if $L_1 = 1.1$ and $L_2 = 0.9$, then K in both cases is equal to 0.1; for $L_1 = 1.2$ and $L_2 = 0.8$, $K = 0.2$, etc. But this equality in the values of K and the equality in the attenuation of the advantage of the method of corrections holds only for the limiting values of K (with $r_{xy} = 1$): for all the remaining course in the variations of r_{xy} these equalities do not hold. We have already noted above that for $L < 1$, this attenuation comes also from the other end, namely: a changeover in favor of the method of corrections is observed at $r_{xy} = 0.50$, and as the inequality $L < 1$ increases, this changeover demands larger and larger values of r_{xy} ; while for $L > 1$, on the contrary, the superiority of the method of corrections begins at $r_{xy} = 0.50$, and the sharper the inequality $L > 1$ becomes, the lower the values of r_{xy} that are required.

Thus in the contrast to the extreme values of K , with $L < 1$ the advantage of the method of corrections "from the other end" is strengthened. We shall now prove the following theorems: (1) if L becomes less than 1, then in connection with this, the advantage of the method of corrections is increased for $r_{xy} < \frac{L+1}{2}$ it remains at the same level for $r_{xy} = \frac{L+1}{2}$, and it is weakened for $r_{xy} > \frac{L+1}{2}$; (2) if L becomes greater than 1, then for all values of r_{xy} the superiority of the method of corrections is diminished. Let us employ formulae (5) and (6). The former expresses the ratio of the error of the difference of the two arithmetic means M_y and M_x to E_y , the error of the mean of the sample study, with $L = 1$. Let $(1)K$ be this ratio. The latter equation (6)

gives the same ratio for $L \pm 1$; let $(L)K$ be this ratio. It was noted in Chapter I that the values $\frac{1}{(1)K^2}$ and $\frac{1}{(L)K^2}$ represent a peculiar kind of "coefficient of useful action" of the method of corrections; they show how many times the number of observations for the independent computation from the precision sampling study must be increased to obtain the same degree of accuracy given by the correction method, or, if the conditions are more favorable for the sampling method, how many times the size of the sample may be decreased for that purpose.

The condition for the strengthening of the superiority of the method of difference, with $L = 1$, is expressed as follows: the "coefficient of useful action" $\frac{1}{(L)K^2}$ must be greater than

$\frac{1}{(1)K^2}$, namely, that $\frac{1}{(1)K^2} \cdot \frac{1}{(L)K^2} > 1$ or $\frac{(1)K^2}{(L)K^2} > 1$ or

$$\frac{(1)K^2}{(L)K^2} < 1, \text{ or } L^2 + 1 - 2Lr_{xy} < 2 - 2r_{xy}.$$

Further: $L^2 - 1 + 2r_{xy}(1-L) < 0$
 $L^2 - 1 + 2r_{xy}(L-1) < 0$
 $(L-1)(L+1-2r_{xy}) < 0$

The multiplier $(L - 1)$ has the minus sign for all values of L , since by hypothesis $L < 1$; therefore if inequality (8) is to hold, the second multiplier $(L + 1 - 2r_{xy})$ must be greater than 0:

As a result we have $r_{xy} < \frac{L+1}{2}$ as the condition which must be satisfied by the numerical values of the correlation coefficients that strengthen the advantage of the method of corrections by comparison with the variant $L = 1$. We obtain the numerical values of the correlation coefficients r_{xy} under which the superiority of the method of corrections is reduced, by solving the inequality $(L)K^2 > (1)K^2$. They must satisfy

the condition $r_{xy} > \frac{L+1}{2}$

Finally, the superiority of the method of corrections remains unchanged if $(L)K^2 = (1)K^2$ which reduces to $r_{xy} = \frac{L+1}{2}$; if $L > 1$, the inequality (8) cannot hold. In this case the multiplier $(L-1)$ is always positive, exactly the same way as the multiplier $(L + 1 - 2r_{xy})$, for $L + 1$ is always greater than 2, and $-2r_{xy}$ can never be more than 2, since $r_{xy} \leq 1$. Thus in this case there is no possibility for the strengthening of the superiority of the method of corrections. The condition for its weakening is expressed by the inequality:

$$(L-1)(L+1-2r_{xy}) > 0$$

If $L - 1$ is positive, $L + 1 - 2r_{xy} > 0$, and as a result we have: $r_{xy} < \frac{L+1}{2}$

Since $L > 1$, then also $\frac{L+1}{2} > 1$.

Consequently all possible numerical values of the coefficients r_{xy} satisfy the condition for the weakening of the advantage of the method of corrections, for correlation coefficients cannot be greater than 1. Thus both theorems have been proved.

If we have different L 's, then the coefficients of useful action for the same value of r_{xy} are likewise different. We put the question: to what value of L does the maximum value of $\frac{1}{(L)K^2}$ correspond, at constant r_{xy} ? Obviously the maximum value of $\frac{1}{(L)K^2}$ occurs at the minimum value of $(L)K^2$ and the matter reduces down to finding a minimum for the function $K^2 = L^2 + 1 - 2Lr_{xy}$ in dependence on the variation of L .

We find the first derivative in L of this function and equate it to zero. As a result we obtain: $2(L - r_{xy}) = 0$.

Knowing that the second derivative has a plus sign for all values of L , we conclude that with $L = r_{xy}$ we have a minimum for $(L)K^2$, and, consequently, the maximum for the coefficient of useful action is $\frac{1}{(L)K^2}$.

Thus, for instance, if the correlation between the readings of the defective source and the readings obtained by the aid of the precise investigation is expressed by a correlation coefficient 0.75, then the most favorable ratio between the variability of one series and the other, $\frac{\sigma_x}{\sigma_y}$ will also be equal to the same value 0.75. (Since $\frac{\sigma_x}{\sigma_y} = \frac{E_x}{E_y} = 1$)

In a word, we obtain the optimum results at a given value r_{xy} when L is equal to the same numerical value; every other value of L , whether larger or smaller than r_{xy} is less favorable. (It follows from the equality $L = r_{xy}$ that if L is greater than 1, there can be no optimum, since r_{xy} cannot be greater than 1. The optimum when $L = 1$ occurs only at the limiting value $r_{xy} = 1$, and for all other values of $r_{xy} < 1$, the optimum value of L is less than 1.)

The schematic table of the "coefficient of useful action" prepared by us vividly confirms the method of corrections by and illustrates the basic conclusions of our work.

It is calculated according to the formula

$$\frac{1}{(L)K^2} = \frac{1}{L^2 + 1 - 2Lr_{xy}}$$

in which the value of

of two variables: (1) the correlation coefficient r_{xy} and (2) L . Numerical values of

in these cases it is more advantageous to employ the independent computations from precision sampling.

The values

$$\frac{1}{(L)K^2} = 1$$

mark the points of "changeover".

Finally, values of

$$\frac{1}{(L)K^2}$$

greater than 1 establish the superiority of the method of corrections and the degree of that superiority.

It can be seen from the data of this table how uninterruptedly the coefficient of useful action grows as r_{xy} grows, at unchanged value of L , and how it varies in dependence on the value of L , at constant r_{xy} , reaching a maximum at $L = r_{xy}$.

In general, the most favorable variants for the method of corrections are those in which L is equal to unity or near it, i.e., cases where σ_y and σ_x differ little from each other. Since the precision sampling study and the complete-coverage defective study both relate to the same criterion (for instance, to the seeded acreage or the yield of a certain territory in a certain year) it is precisely with these variants of L that we must deal in practice. (In the course of my researches I have been able to convince myself of this. Although in investigating the extent of the seeded acreage omitted from record, and studying the corrections to the materials of the autumn questionnaires on the yields for prior years, I was confronted by the differences between the control sources (surveys of seeded acreage, test reaping and thrashing), and mass observations ranging from 5 to 35%, the standard deviations σ_y and σ_x of the precise and defective observations almost coincided. (The circumstance that the errors of observation are fairly low in comparison to the real variability of the criterion itself is a point of substantial importance. [See footnote on page 207.]

III

In the present article we have examined one method of correcting a solid-coverage but defective source. This is the method of superimposing the absolute amounts of the corrections, as disclosed by comparing the results of both sources for n sampled units ($M_y - M_x + x_0$).

The second variant differs from the first by the superimposition, together with the absolute correction, of a relative correction as well. Finally, the third method, which is the most in-

interesting and complete one, is the use of n sample observations to set up a regression equation $y = a + bx$ (y is the precision sampling, and x the observations taken from the solid coverage but defective study). Substituting in this equation x_0 , the mean value of x for the total population, we determine the most probable value of the phenomenon with which we are concerned: $y_0 = a + bx_0$. The approximate (but only the approximate) error y_0 is expressed by the formula

$$\frac{s_y}{\sqrt{n}} \sqrt{1 - r_{xy}^2}$$

It is entirely clear that in this case the coefficient of useful action of the method of corrections

$$\frac{1}{2K^2} = \frac{1}{1 - r_{xy}^2}$$

The following schematic table gives its numerical value at different values of r_{xy} . pK characterizes the application of the regression equation in comparison with the calculation, directly based on the results of the sampling study, while $1K$ is the analogous index for the method of corrections.

[see Table on following page]

As may be seen from the formula and the schematic table constructed on the basis of it, the method of correction by the aid of the regression equation "in principle" for all values of r_{xy} not zero, yields better results than direct computation for n sample units of the precision study. The tangible dimensions of this superiority are disclosed by the correlation coefficients near 0.60 when the method of the regression equation makes it possible to cut the number of precision sampling measurements by one and a half times.

r_{xy}	$\frac{1}{r^2}$	$\frac{1}{1/r^2}$
0	1.0	
0.25	1.07	
0.50	1.33	1.00
0.60	1.56	1.25
0.70	1.96	1.67
0.75	2.29	2.00
0.80	2.78	2.50
0.85	3.60	3.33
0.90	5.26	5.00
0.95	10.26	10.0
0.975	20.24	20.0

On comparing these results with the variants considered in detail, we must give preference to the method of the regression equation, firstly because the "coefficient of useful action" that belongs to it for all values of r_{xy} is greater than unity; secondly because for the same correlation coefficients it gives smaller errors; and thirdly it does not depend on the ratio of σ_y to σ_x (This is not entirely correct. A sampling measurement study performed by more accurate instruments has a smaller error of observation (which is reflected in the value of L ; cf. footnote on page 307)). That part of the dispersion which reduces the correlation between x and y is diminished thereby, while the correlation coefficient r_{xy} is increased, which is reflected in the values of K . (Editor's note)).

We note that all methods of correcting the defective solid-coverage investigation are based on two propositions.

The first proposition: we use the difference $(M_x - x_0)$ between the mean values of the investigation to be corrected, for n sample observations, and those of the entire population.

The second proposition: such use is effective only when there is a definitive correlation between the defective series x and the accurate series y , established for n sample observations.

Actually, for the first method the most probable value of the phenomenon under study is defined by the formula:

$$y_0 = M_y - M_x + x_0 = M_y - (M_x - x_0)$$

This expression shows that the correction of the results of solid-coverage defective observation is transformed into the correction of the results of n precise sample observations (M_y) by subtraction of the algebraic difference $M_x - x_0$.

The same thing also takes place with correction by the regression equation method:

$$y_0 = a + bx_0 = M_y - \frac{\sigma_y}{\sigma_x} r_{xy} M_x + \frac{\sigma_y}{\sigma_x} r_{xy} x_0 = M_y - \frac{\sigma_y}{\sigma_x} r_{xy} (M_x - x_0)$$

but this variant is corrected by subtraction of the same difference $(M_x - x_0)$, multiplied in this case -- in contrast to the first method -- by a certain multiplier.

$$\left(\frac{\sigma_y}{\sigma_x} r_{xy} \right)$$

We direct our attention first to the element common to both methods, i.e., to the difference $M_x - x_0$.

If M_x is greater than x_0 , then this difference serves as an indicator that the sample collection of n observations is faulty in that x violates representativeness on the side of exaggeration; and since the series y and x are intercorrelated, it is probably the value of M_y is also at fault in the same way -- in any case if it is over 0.5. For this reason M_y must be reduced by a certain value in order to bring it closer to the truth. The first method reduces it by the value $(M_x - x_0)$, while the second method reduces by the same value, multiplied by the multiplier

$$\frac{\sigma_y}{\sigma_x} r_{xy}$$

In the contrary case, if M_x is less than x_0 , then the violation by x of the representative character of the sample collection makes it necessary to correct M_y , on the basis of the same considerations, by means of a certain increase in it: in the first method, by $(x_0 - M_x)$, and in the regression equation variant, by the value of

$$r_{xy} \frac{\sigma_y}{\sigma_x} (x_0 - M_x)$$

Thus both variants, having the same value $(M_x - x_0)$ as a starting point, use it in different ways.

Which is more accurate?

Mathematical treatment shows that preference should be given to the regression equation method: the errors in the computed values of y_0 are less in this case (cf. the table of coefficients of useful action). But then it could not be otherwise.

The method of adding the divergence $(x_0 - M_x)$ to the value of M_y leads, as we have seen, to the correction of the mean value of n precise sample measurements (M_y) by subtracting from it the algebraic difference $(M_x - x_0)$. Starting from the existence of a correlation between x and y , this method, however, (1) does not take the closeness of this correlation into account, and (2) does not make use of the ratio between the respective variabilities of these series -- the ratio of their standard deviations σ_y and σ_x -- for the translation of the difference $(M_x - x_0)$ to the scale of the measurements $(M_y - y_0)$. (y_0 is the mean value for the total population of precise observations if these had completely covered the field and had not been confined to n sampling observations.) The indices r_{xy} , σ_y and σ_x , are used to calculate the errors, defining the values of y_0 by this method, but these indices are absent from the process of correction.

None of these faults occur in the regression equation variant.

$\frac{\sigma_y}{\sigma_x} r_{xy}$, which is the multiplier for $(M_x - x_0)$, translates the scale $(M_x - x_0)$ to the scale $(M_y - y_0)$, by the aid of $\frac{\sigma_y}{\sigma_x}$, and thus transforms the operations on M_y into operations on values uniform with respect to M_y , while with the aid of r_{xy} it makes to M_y a correction proportional to the closeness of the connection between the series x and y .

The transition from the less perfect method to the more accurate one involves no difficulties, since the values σ_y , σ_x , r_{xy} , which are necessary for the regression equation, must all be calculated in any case to determine the errors of the first variant.

However, even this coarser first method yields, as we have seen, better results than the direct use of n precise sampling measurements without considering the indices of the defective mass investigation -- provided that the correlation coefficients are high enough and that certain ratios obtain between σ_y and σ_x . These conditions are often met in practice. It is obvious that in these cases the defective mass study method is distinguished by even greater coarseness and inaccuracy. The divergence between the results of the calculation by the two methods is given by the following formula:

$$y_1 - y_2 = (M_x - x_0) \left(1 - \frac{\sigma_y}{\sigma_x} r_{xy} \right)$$

It appears from this formula that: (1) with $(M_x - x_0)$ constant, the extent of the non-coincidence is determined by the deviation of unity of the coefficient of regression

$$\left(1 - \frac{\sigma_y}{\sigma_x} r_{xy} \right); \text{ when } \frac{\sigma_y}{\sigma_x} r_{xy} = 1, \text{ then}$$

both standards give identical results, for y_1 and y_2 are both calculated by the same formula.

$$y = M_y - (M_x - x_0)^2$$

(2) with $\frac{\sigma_y}{\sigma_x} r_{xy}$ constant, the difference in the results of the computations is the greater, the greater the violation of the representative character of the sampling observations is (i.e., the greater the value of $M_x - x_0$). (x_0 is the mean value for the total

population of precise observations if these had completely covered the

only field and had not been confined to n sampling observations. $\frac{\sigma_y}{\sigma_x} r_{xy} = 1$ is
We dwelled on this case in Chapter II. We can now understand why

it yields the most favorable results: it coincides with the regression equation variant. The errors, of course, are identical; if $L = r_{xy}$, the expression

$$\frac{\sigma_y}{\sqrt{n}} \sqrt{L^2 + 1 - 2Lr_{xy}} = \frac{\sigma_y}{\sqrt{n}} \sqrt{1 - r_{xy}^2}$$

We shall still dwell briefly on the method of relative corrections mentioned in Chapter III, namely

$$\frac{M_y}{M_x} X_0$$

This method represents another form of the first variant, as will be seen from the following identity:

$$\frac{M_y}{M_x} X_0 = M_y - \frac{M_y}{M_x} (M_x - X_0)$$

Correcting the value of M_y , this form starts with the difference $M_x - x_0$ just as do the variants we have considered. In failing to take account of the closeness of the relationship between the series x and y , it shares the defects of the method of superposing absolute corrections. But in contrast to the latter, the variant of relative corrections translates the scale of the difference $M_x - x_0$ into the scale of the values of M_y by using the multiplier $\frac{M_y}{M_x}$. This is, of course, less perfect a method than using the ratio of σ_y and σ_x , but it is still an improvement, since we have reason to assume that the values of the standard deviations (σ_y and σ_x) are to some extent dependent on the values of the series y and x . (For example, if $M_y > M_x$, we have reason to believe that $\sigma_y > \sigma_x$, and conversely.)

From everything we have said, and from the formulae, it is easy to convince oneself that if M_x and x_0 completely coincide,

the results of the computations by all the correction variants will likewise coincide and become equal to M_y , i.e., to the result of the direct determination of the phenomenon that interests us by means of n precise sampling measurements.

At first glance, there would seem to be no advantage, in this case, of carrying out the computation of the elements necessary for the regression equation, and we could be satisfied with the results yielded by the use of n precise sampling measurements. But this would not be correct, since it is important for us to know not only the most probable value of the phenomenon that interests us (M_y in the given case) but also the amount of its inaccuracy; but, where there is a correlation between x and y , this error is defined by the formula

$$\frac{\sigma_y}{\sqrt{n}} \sqrt{1 - r_{xy}^2}$$

and not by the formula usually applied to random sampling selections

$$\left(\frac{\sigma_y}{\sqrt{n}} \right)$$

We conclude our sketch with the following observation.

All of the variants discussed refer to the form of random selection of n sample units from the total population (for instance, by 11 lot).

The matter becomes more complicated when we use what is termed mechanical sampling, which presupposes the existence of a list of all objects comprising the population, from which list n units are drawn, at stated intervals, to make up the sample collection.

May this method of sampling be considered completely identical to random selection? A positive or negative answer to this question depends on the base on which the list of the population has been compiled. If it has been compiled in random order, with respect to the criterion that interests us, so that this criterion will not react even "correlatively" to the position of the units in the list, then the analogy with the type of sampling study already considered appears perfect. In this case the selection of units at definite intervals is an operation equivalent to drawing lots. But if the size of the phenomenon interesting us is "correlatively" connected to position on the list, then the conditions for random sampling are not entirely realized. The selection of farm units for yield determination on the basis of a list drawn up in geographical order may serve as an example. To the extent that the yield of a farm unit depends on its geographical location (soil conditions, climatic and meteorological factors in the given year), a correlation arises between yield and its position (the number it occupies on the list).

It is easy to convince oneself that mechanical sampling in this variant gives better results than random selection, since we are "aiming" to obtain in the sample collection a distribution curve closely resembling the distribution curve of the original population. Consequently the error in the value of M_y is lower than with random selection. The question is further complicated by the fact that the transition from σ_y to the error M_y with the help of the multiplier $\frac{1}{\sqrt{n}}$ can hardly be applied to mechanical sampling from a list distinguished by the correlative relationship noted above.

From a manuscript completed in 1932

CONTENTS

	Page
Preface	3
Sketch of yield dynamics during the period 1883-1915	9
On the question of forecasting the character of the spring and summer weather	42
The influence of meteorological factors in the fluctuations of yield in the arid Povolzh'ye at varying levels of cultural practice	119
The influence of meteorological factors on the fluctuations in the yield of spring and winter wheat in the chernozem belt (from materials of the agricultural experiment stations)	213
On the question of the determination of the expected yield (from materials of Moscow and Petrograd guberniyas)	244
On the question of the methods of utilization of sample surveys	298

Technical editor V.T. Krashain

A 07427 Signed for printing 4 August 1948 Contents 20 quires

Accounting-publishing quires 23.5 Characters per quire 48144

Edition 6000 copies (1-3000) Order 273

Printing plant of GOSPLANIZDAY imeni Vorovskiy, Kaluga